

Supplementary Methods

Sampling sites and sample preparation

The samples used in this study were obtained from 9 sampling sites (**Fig. 1, upper panel**) and each sample (except for HSM and HOT) was separated into two parts, a "Single Cell" aliquot and "DNA isolation" aliquot. The "Single Cell" aliquot was cryopreserved in a betaine solution or glycerol for single-cell genomics (see detailed sampling procedures below). The "DNA isolation" aliquots were stored at -80 °C, transported to the laboratory, and used to assess the SSU rRNA community profile (see details below). Physicochemical conditions were recorded for each sampling site (**Supplementary Table 12**).

Detailed sampling procedures

Sakinaw Lake (SAK). Samples were collected on January 5th 2010 in brackish water in 120 m depth in Sakinaw Lake (49°40'30"N, 124°2'2.4"W). For the "Single Cell" aliquot 1 ml water was cryopreserved in 143 µl 48% betaine (anhydrous, Fisher #AC20424-1000) on board of the sampling vessel shortly after bringing the water up. Samples were then directly frozen on dry ice and kept at -80 °C until further processing.

Etoliko Lagoon (ETL). Sediment from the deepest point of the Etoliko lagoon was collected on January 19th 2009 (38°28'59.54"N, 21°19'17.44"E) and placed in an anaerobic stainless steel jar. The sample was transferred immediately to the lab and placed in an anaerobic chamber (glove box) where sample preparation took place. In detail, 75 g of the sediment was weighed and transferred to a 2 l flask containing 1 l of benthic water from the Etoliko lagoon (filter-sterilized). The sample was gently agitated for 30 min and then it was filtered through: (a) sterile hydrophilic cotton to remove large aggregates, (b) 20 µm Millipore nylon filters and (c) 3 µm glass fiber filters. Finally, 1 ml of the filtered sample was placed in a cryovial tube with 143 µl of 48% betaine and was frozen in liquid nitrogen as the "Single Cell" aliquot.

Terephthalate degrading reactor (TA). The seeding sludge for the reactor was obtained from the mesophilic UASB reactors treating dimethylterephthalate (Temex and Petrocel, Mexico; 25°40'18"N, 100°20'30"W). The TA reactor was operated at 45-50 °C, fed with terephthalate disodium at 3500 mg/l and 800~1000 ml/day, with recirculation ratio around 66. The total volume of the TA reactor was 880 ml (working volume 841 ml) and 51 siporax biofilm carrier (around 39 ml) are stacked up from the middle of the reactor. Two samples were obtained and later combined for single-cell genomics. Sample 1 was a "biofilm sample" scratched from attachment media (siporax biofilm carrier) with a cotton stick, sonicated in a water bath for 10 min, and shaken by hand for 5 min to loosen the tight biofilm aggregate. Sample 2 was a "sludge sample" obtained from the bottom of the reactor by sampling 50 ml of reactor liquid, which was centrifuged to form a cell pellet. The liquid volume was adjusted to 15 ml in the 50 ml centrifuge tube, 0.1 g glass beads were added into the tube, and the pellet was ground by hand for 10 min to loosen the tight aggregates. After a resting period of 3 min the upper suspension was poured into a new tube. The cell density of both samples was estimated to be 10⁵ to 10⁶ cells/ml. Finally samples for the "Single Cell" aliquot were suspended in 15% glycerol (final concentration), and stored at -80 °C.

Gulf of Maine (GOM), Hawaii Ocean Time-series (HOT), Tropical Gyre Atlantic (TG) and Homestake Mine (HSM). Water samples were collected from surface and mesopelagic depths using Niskin bottles during cruises in the South Atlantic (TG; 27 November 2007, 800 m, Cruise KN192-5 station 11, 12°29'41.4"S, 4°59'55.2"W) and North Pacific (HOT; 9 September 2009, 25 m and 770 m; Cruise 215, station ALOHA; 22°45'N, 158°00'W). The Homestake mine (HMS) sample was collected at the Homestake Mine, Lead, South Dakota (44°20'46"N, 103°45'39"W), on April 20, 2008. The collected aquatic sample is a groundwater seep from the ceiling of a mine drift at about 100 m depth. Replicate

samples of 1 ml aliquots of water of all four aforementioned sampling sites were cryopreserved with 6% glycine betaine (Sigma) and stored at -80 °C for the "Single Cell" aliquot.

East Pacific Rise (EPR). Water samples were collected in 2503 m depth at the diffuse flow deep-sea hydrothermal vent site "CrabSpa" (9°50'12"N, 104°17'30"W) on the East Pacific Rise. Replicate samples of 1 ml aliquots of water were cryopreserved with 6% glycine betaine (Sigma) or 15% glycerol and stored at -80 °C for the "Single Cell" aliquot.

Great Boiling Spring (GBS). Samples were collected at the Great Boiling Spring (40°39'41.16"N, 119°21'58.5"W). Two samples, one from the top layer (~1 cm) of the sediment in the main spring pool (~85 °C) and one from the top layer (~1 cm) of the sediment near the spring outflow (~75-80 °C) were collected and later combined for the analysis. Cells were separated from sediment by centrifugation over a Nycodenz cushion and the "Single Cell" aliquot was preserved in betaine (1 ml sample and 143 µl of 48% betaine) and stored at -80 °C.

Cell sorting, lysis, single-cell amplification, and SSU rRNA amplicon sequencing

Fluorescence Activated Cell Sorting was performed with a Cytosort Influx Cell Sorter (BD Biosciences) into 384-well plates containing 0.9 µl of UV-treated TE. The cells were stained with SYBR Green I (Invitrogen) and illuminated by a 488 nm laser (Coherent Inc.). The sorting window was based on size determined by side scatter and green fluorescence (531/40 bp filter). For each plate, single cells were sorted into 24 columns, including three columns of negative controls and one column of positive controls with 100 and 10 cells per well. The sort purity was determined by depositing a predefined number of 2 µm fluorescent beads on glass slides and the results were checked via epifluorescence microscopy to verify the presence of the correct number of beads. Sorted cells were lysed for 10 min at room temperature using alkaline solution from the Repli-G UltraFast Mini Kit (Qiagen) according to manufacturer's instructions. After neutralization, the samples were amplified using the Replphi Phi29 reagents (Epicentre). Each 15 µl reaction contained Phi29 Reaction Buffer (1X final concentration), 50 µM random hexamers with phosphorothioate modification of the two 3'-terminal nucleotides (IDT) (Dean et al, 2002), 0.4 mM dNTP, 5% DMSO (Sigma), 10 mM DTT (Sigma), 100 U Phi29 and 0.5 µM Syto 13 (Invitrogen). A mastermix of MDA reagents minus the Syto 13 (degrades when exposed to UV) sufficient for a 384-well plate was assembled and then aliquoted into four Eppendorf Safe-Lock 1.5 ml clear microcentrifuge tubes. The tubes of mastermix were UV treated on ice in the Stratalinker 2400 UV Crosslinker (Stratagene) at 254 nm for 30 to 90 min (Woyke et al., 2011). These represent a UV doses of 0, 5.7 to 17.1 J/cm², respectively, when measuring inside the eppendorf tubes at the 4 cm distance to the light bulb. Following UV treatment the mastermix tubes were combined, Syto 13 was added, and the mixtures were aliquoted into each well. The MDA reactions were run in real time on the Roche LightCycler[®] 480 II real-time thermal cycler for 16 h at 30 °C. Cell sorting, lysis, whole genome amplification and SSU rRNA PCR for a subset of samples was performed at the Bigelow Laboratory Single Cell Genomics Center (http://www.bigelow.org/index.php/research/facilities/single_cell_genomics_center/) according to the procedure described in Swan et al. 2011.

For the SSU rRNA screening the MDA products were diluted 20 to 50-fold in TE buffer and 1 µl aliquots of diluted MDA product served as the template DNA in 10 µl final volume real-time PCR screens. All PCR reactions were performed using KAPA[®] SYBR FAST qPCR kit (KAPA) and the Roche LightCycler[®] 480 II real-time thermal cycler. PCR amplification of SSU rRNA was carried out using the primers: Forward 926wF (GAACTYAAAKGAATTGRCGG) and reverse 1392R (ACGGGCGGTGTGTRC). In addition full length SSU amplification was performed using the primers: 27F (AGAGTTTGATCCTGGCTCAG) for Bacteria, 4aF (TCCGGTTGATCCTGCCRG) for Archaea, together with the universal reverse primer 1391R (GACGGGCGGTGWGTRCA). M13 sequencing primers FW (GTTTCCAGTCACGACGTTGTA) and RV (AGGAAACAGCTATGACCAT) were added to the 5' ends of each target primer pair to aid direct sequencing of PCR products. All PCR reactions

were run for 35 cycles at 60 °C annealing temperature, followed by melting curve analysis performed as follows: 95 °C for 5 s, 65 °C for 1 min, and a continuous temperature ramp (0.11 °C/s) from 65-97 °C. Real-time PCR kinetics and amplicon melting curves served as proxies for detecting SAGs positive for SSU rRNA genes and single-cell wells with positive PCR amplification were cherry-picked and Sanger sequenced using the 96-capillary ABI 3730xl DNA Analyzer.

SSU rRNA gene pyrotag analysis of microbial communities

The “DNA isolation” aliquot of the sample was used for DNA extraction utilizing a FastDNA Spin Kit for Soil (MP Biomedicals, Solon, OH, USA). DNA was precipitated with 70% ethanol, resuspended in 0.5X TE and quantified using a Nanodrop 1000 (Thermo Scientific, Waltham, MA, USA). Based on this quantification, a quantity of the extracted DNA equivalent to 1,000,000 16S rRNA gene copies was used as template for a preparative PCR to generate adapted and barcoded 16S rRNA gene amplicons for pyrosequencing. Primers 926wF and 1392R (see above) modified with Titanium pyrosequencing adaptors A (926F) and B (1391R) and a 10-nucleotide barcode were used to amplify SSU rRNA genes from Bacteria and Archaea. The reactions were performed in 50 µl total volume using Platinum HiFi Master Mix (Invitrogen, Carlsbad, CA) on an MJ PTC-100 thermocycler (MJ research/Bio-Rad Laboratories, Hercules, CA) with the following cycling conditions: 94 °C for 5 min; 24 cycles of 92 °C for 20 s, 50 °C for 30 s, 65 °C for 60 s, 75 °C for 60 s, 70 °C for 60 s; and a final 10 min extension step at 70 °C. Products were purified using the Qiaquick PCR purification kit (Qiagen, Valencia, CA), followed by purification by Solid Phase Reversible Immobilization and size-selection step on calibrated Agentcourt AMPure magnetic beads (Beckman Coulter Genomics, Danvers, MA) according to the Roche/454 method. Purified libraries were quantified using the previously described digital PCR method (White et al., 2009 [353]), and sequencing of the 16S rRNA gene amplicons was performed on a Roche 454 Genome Sequencer FLX System using Titanium chemistry according to the manufacturer’s protocols (454 Life Sciences, Branford, CT). Emulsion PCR was carried out with DNA : bead ratios between 0.08:1 and 0.3:1.

Analysis of the resulting sequence data was performed using PyroTagger (<http://pyrotagger.jgi-psf.org>; Kunin and Hugenholtz, 2010). A quality filter of 10% was used for all data. Sequences were trimmed from the 3’ end to 250 nucleotides in length and OTUs (grouped at 97% sequence identity) were determined using the pyroclust algorithm (Kunin and Hugenholtz, 2010). PyroTagger assignments for clusters containing 1% or more reads were manually checked and corrected.

SAG assembly

The draft genome of all but 13 SAGs was generated at the DOE Joint genome Institute (JGI) using the Illumina technology. An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform. All general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov/>. Raw Illumina sequence data were filtered for known Illumina sequencing and library preparation artifacts and then screened and trimmed according to the k-mers present in the dataset. High-depth k-mers, presumably derived from MDA amplification bias, cause problems in the assembly, especially if the k-mer depth varies in orders of magnitude for different regions of the genome. Reads representing highly abundant k-mers were removed such that no k-mers with a coverage of more than 30x were present after filtering. Reads with an average kmer depth of less than 2x were removed. The following steps were then performed for assembly: (1) filtered Illumina reads were assembled using Velvet version 1.1.04 (Zerbino and Birney, 2008). The VelvetOptimiser script (version 2.1.7) was used with default optimization functions (n50 for k-mer choice, total number of base pairs in large contigs for cov_cutoff optimization). (2) 1-3 kbp simulated paired end reads were created from Velvet contigs using the wgsim software. (3) the normalized Illumina reads were assembled together with simulated read pairs using Allpaths-LG (version 41043) (Gnerre and MacCallum, 2011). Parameters for assembly steps were: 1) VelvetOptimiser (--v --s 51 --e 71 --i 4 --t 1 --o "-ins_length 250 -min_contig_lgth 500") 2) wgsim (-e 0 -l 100 -2 100 -r 0 -R 0 -X 0) 3) Allpaths-LG (prepareAllpathsParams: PHRED_64=1 PLOIDY=1 FRAG_COVERAGE=125 JUMP_COVERAGE=25 LONG_JUMP_COV=50, runAllpathsParams: THREADS=8 RUN=std_pairs TARGETS=standard VAPI_WARN_ONLY=True OVERWRITE=True). For the remaining 13 SAGs the draft genomes were generated at the JGI using a Roche 454 Genome Sequencer FLX System using Titanium chemistry according to the manufacturer’s protocols (454 Life Sciences, Branford, CT). The hybrid 454/Illumina assemblies steps (1)-(3) were identical. Next, (4) Allpaths contigs larger than 1 kbp were shredded into 1 kbp pieces with 200 bp overlaps. (5) Lastly, the Allpaths shreds and raw 454 pyrosequence reads were assembled using the 454

Newbler assembler version 2.4 (Roche). All assemblies are available on the Microbial Dark Matter project website (<http://genome.jgi.doe.gov/MDM/>)

Genome annotation

Genes were identified using Prodigal (Hyatt et al., 2010) and the predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScan-SE tool (Lowe et al., 1997) was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA (Pruesse et al., 2007). Other non-coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL (<http://infernal.janelia.org>). Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG) platform developed by the DOE Joint Genome Institute, Walnut Creek, CA, USA (Markowitz et al., 2009).

SAG QC analysis

The first step of single amplified genomes (SAGs) QC was done on annotated SAGs in IMG. Scaffolds of possible contaminants were identified manually for all 201 SAGs using IMG functions including (1) phylogenetic distribution of genes (hits > 90% identity), (2) GC content graph of scaffolds (outliers outside of a +/- 10% range), and (3) Kmer Frequency Analysis (outliers of main cloud using two different settings: fragment window 5000 bp, fragment step 500, oligomer size 4, minimum variation 10; and fragment window 2000 bp, fragment step 500, oligomer size 4, minimum variation 10). Identified scaffolds of interest were blasted against NCBI (BLASTP), hits of genes above 30% identity (except for ribosomal proteins which had a 50% identity cutoff) were automatically analyzed and scaffolds were flagged as contamination if the majority of genes (>50% of genes per scaffold) had hits to the same phylum - or class in case of proteobacterial hits. All flagged scaffolds were manually evaluated and those identified as contamination were binned in IMG into a 'contamination bin' and the remaining scaffold into a 'clean bin' for each SAG. The 'clean bin' was exported for the second step of QC comprising redundancy filtering, cross-contamination filtering and size cutoff. Redundancy filtering has been performed using BLASTN in order to identify repetitive regions in an assembly. Sequence regions with more than 99.5% identity over at least 2000 bp were excised and collapsed into one representative contig. Cross-contamination between samples was analyzed by searching all assemblies against each other using BLASTN. Sequence regions that occurred in assemblies of different OTUs with more than 99.5% identity over at least 5000 bp were excised and removed from the smaller contigs. Finally contigs smaller than 2000 bp were discarded from the assemblies. All quality controlled SAGs were re-submitted to IMG.

Combined assemblies

SAGs were selected for combined assembly if their 16S rRNA genes showed an identity greater than 99%, if they origin from the same sampling site, and if the average nucleotide identity (ANI) of their genomes was greater than 97%, which is based on the proposal that genomes showing more than 95% ANI should belong to the same species (Konstantinidis et al., 2006). ANI calculations were performed with JSpecies (Richter & Rosselló-Móra 2009; ANIb, standard parameters) based on at least 10 kb DNA of greater than 70% BLAST hit identity in order to be counted as valid results. Following these selection criteria, 21 groups of SAGs were established, comprising 122 SAGs from 10 different candidate phyla. The first step involved mapping the raw reads to the pooled SAG assemblies for each group. The mapped reads were fed into the assembly pipeline as described above (SAG assembly).

Genome completeness estimation

Genome size and completeness was estimated using two conserved single copy gene (CSCG) sets that have been determined from all bacterial (n=1516) and all archaeal (n=111) finished genome sequences in the IMG database (Markowitz, 2012). The sets (**Supplementary Table 13**) consist of 139 CSCGs (bacterial) and 162 CSCGs (archaeal) that were found to occur only once in at least 90% of all genomes by analysis of an abundance matrix based on hits to the protein family (Pfam) database (Punta et al., 2012).

HMMs of the identified Pfams were used to search both, all SAG assemblies and all combined assemblies by means of the HMMER3 software (Eddy, 2011). Resulting best hits above pre-calculated cutoffs were counted and the completeness was estimated as the ratio of found CSCGs to total CSCGs in the set after normalization to 90%. The estimated complete genome size was calculated by division of the estimated genome coverage by the total assembly size.

Phylogenetic analysis

Genome assemblies were scanned for homologs of a set of 38 universally conserved single-copy proteins in the bacteria and archaea (Wu et al., in prep.). The alignment procedure can be briefly summarized as follows. First, a six-frame translated search of each genome was conducted against a database of phylogenetically diverse representatives of the 38 proteins using the LAST software (Kielbasa et al., 2011). Frameshift detection in query genomes was enabled with a penalty value of 15 to cope with any indel errors in assemblies. Regions of the query genomes covered by a single high scoring alignment to a single marker gene were extracted and added to a multiple sequence alignment of the marker gene using hmalign from HMMER 3.0 (Eddy, 2010). Short alignments containing fewer than 20 amino acid residues were discarded. Where possible, homologs of these proteins from eukaryal reference genomes were included in the alignment. The individual alignments of all 38 proteins were concatenated to form an alignment "supermatrix" for phylogenetic inference. This process is automated in the open source software phylsift (Darling et al., in prep; <http://phylsift.wordpress.com/>).

To evaluate the robustness of our protein trees we used 15 different out-group taxon configurations (alignments (a) to (o), see **Supplementary Table 4a**). We generated three concatenated alignments of up to 38 single copy genes including (a) all bacterial and archaeal species, (b) all Bacteria, and (c) all Archaea. Alignments of solely bacterial or archaeal taxa were used to construct unrooted trees for each domain to address possible long branch attraction of bacterial phyla to the archaeal outgroup and vice versa. In order to test for a possible bias introduced by short alignments, we further subdivided each of the three alignments based on the alignment length and the number of marker genes present. The alignment length subdivisions included taxa with full alignment lengths (over 4000 amino acids; d, e, f), and taxa with sequence length over 2000 amino acids (g, h, i). The subdivisions based on the number of present marker genes included all taxa with greater than or equal to 30 (j, k, l), and 20 marker genes (m, n, o; **Supplementary Table 4a**). Phylogenetic inference methods used were the maximum likelihood based FastTree2 (Price et al., 2010), RAxML-Light (Stamatakis et al., 2012), and RAxML version 7.3.5 (Stamatakis, 2006). The amino acid evolution models and distance corrections used are provided in **Supplementary Table 4a**. Two representative trees (tree nb 2 and nb 3; **Supplementary Table 4**) were selected, phyla were grouped into clades and manually corrected in ARB, newick trees were exported from ARB and beautified with iTOL (Letunic and Bork, 2011). For the calculation of the phylogenetic tree investigating the position of the Patescibacteria within the Terrabacteria, we used a custom marker set of 83 bacteria specific markers (**Supplementary Table 14**), and calculated a maximum likelihood tree (fasttree, JTT, CAT).

To ensure that our phylogenetic marker genes yield comparable phylogenetic signals we calculated individual maximum likelihood gene trees (RAxML version 7.3.5 (Stamatakis, 2006) including all organisms sharing a core of 31 marker genes. The quantitative comparison of the resulting phylogenetic trees was performed by calculating the pairwise geodesic distance (Owen and Provan, 2009). The geodesic distance incorporates branch lengths as well as the tree topology and provides a powerful tool for studying and comparing phylogenetic trees (Owen and Provan, 2009). We show the frequency of pairwise geodesic distances (**Supplementary Fig. 7a**) and the hierarchical clustering of pairwise distances calculated with hclust in the R software environment (<http://www.r-project.org/>, **Supplementary Fig. 7b**).

Phylogenetic inference of SSU rRNA genes

The de-replicated (at 99% nt identity) SSU rRNA database (version 12-10-2012) was obtained from the greengenes mirror site (ftp://greengenes.microbio.me/greengenes_release; McDonald et al., 2012), further de-replicated to 97% nt identity, loaded into ARB (Ludwig et al., 2004) and manually curated. De-replication was achieved by all against all aligned sequence comparisons. All matching base positions for each pair of sequences was summed and divided by the maximum length of both sequences (end-gap free). If the resulting ratio was above 97% nt identity, the query read was discarded. To obtain the most suitable SSU RNA sequences for the single-cell amplified genomes (SAGs), we compared quality checked (Q20) DNA sequences obtained from short and full length PCR amplifications (see paragraph “Cell sorting, lysis, single-cell amplification, and SSU rRNA amplicon sequencing”) as well as SSU rRNA genes derived from the shotgun sequencing assembly which ranged in length from 388 to 1566 bp. Following an identity check (97% minimum ID) the longest sequence for each SAG was selected, aligned and added to the 97% ID database, resulting in a total of 4948 taxa (481 Archaea, 4467 Bacteria). Full length sequences over 1000 bp were exported from ARB using a termini filter to restrict the alignments to *E. coli* positions 25 to 1250. Maximum likelihood trees were calculated from the masked alignments using the Generalized Time-Reversible model with CAT approximation with 20 rate categories in FastTree2 (Price et al., 2010). Trees were re-imported into ARB, and the remaining short sequences were added using the “ARB parsimony quick add marked” function with a positional variability by parsimony filter. The resulting tree was rooted, beautified and grouped in ARB (**Supplementary Fig. 8**).

Phylogenetic inference of single protein-coding genes

Homologs of genes of interest identified in SAG genomes (**Supplementary Figs. 14–18**) were obtained and downloaded from IMG using the ‘Find IMG homologs’ function and also obtained from Genbank by BLASTP of the target gene. The obtained homolog datasets were comparatively aligned against their respective COGs (**Supplementary Table 15**) in NCBI’s Conserved Domain Database (CDD; Marchler-Bauer et al., 2013) using MAFFT version 6 (Kato and Frith, 2012). The resulting protein alignments in fasta format were imported into ARB (Ludwig et al., 2004) and positions that could not be unambiguously aligned were removed by generating a termini filter for the alignment. Maximum likelihood trees were calculated from the masked alignments using the JTT model and CAT approximation with 20 rate categories in FastTree2 (Price et al., 2010). The resulting tree was rooted, beautified and grouped in ARB.

Determination of the amino acid found to be encoded by the UGA stop-codon

The gene prediction software Prodigal (Hyatt et al., 2010) was used to predict two sets of open reading frames (ORFs). The first set was derived with the standard translation table 11 (UGA→Stop), whereas the second set was calculated based on translation table 4 (UGA→Trp). ORF length statistics between the two sets were then calculated. The DNA sequence for each ORF of the second set was then translated with translation table 11 in order to yield a third set, containing stop-codon-markers at the respective amino acid positions. Each of these amino acid sequences was then searched against NCBI’s non-redundant protein sequence database (nr) and the alignment to the best hit was examined for the aligned amino acids that corresponded to the stop-codon-markers. Statistics were calculated from all the derived amino acids found to be aligned to the stop-codon-markers. Final gene calling was performed with a custom translation table (UGA→Gly).

Phylogenetic Anchoring

Effective phylogenetic anchoring of metagenomic reads was determined with the help of MEGAN4 software (Huson et al., 2011). Candidate phyla that were not present in the NCBI taxonomy database were added to the database prior to any MEGAN classification. Two runs of MEGAN were performed for a random subset of 10,000 reads for each of the 893 analyzed metagenomes. The first run was based on BLASTX (`blastx -num_alignments 50 -num_descriptions 50 -soft_masking true -num_threads 4 -query <query> -out <out.blast> -db <nr.db>` with a default E value cutoff of 10) results against NCBI’s non-redundant (nr) database without modification. For the second run, all finished SAG genomes of the current

study were added to the nr database and the same random subsamples were blasted against this modified database. A paired Student's t-test of the top 19 recruiting metagenomes (see **Fig. 4**) resulted in significantly ($p=0.00024$) more reads that were assigned by MEGAN beyond domain level for the second dataset. Both resulting datasets were analyzed in regards to the percentage of total novel assigned reads ($n=35,728$) to the candidate phyla after the addition of the SAG dataset (Marinimicrobia [39.95%], Aminicenantes [11.12%], Cloacimonetes [7.28%], Fervidibacteria [6.32%], Omnitrophica [5.79%], Parcubacteria [4.78%], Latescibacteria [4.43%], Microgenomates [4.18%], Atribacteria [3.99%], Aerophobetes [1.94%], BRC1 [1.79%], Hydrogenogenetes [1.31%], Thaumarchaeota [1.25%], Aigarchaeota [1.01%], Nanoarchaeota [0.99%], Armatimonadetes [0.53%], Calescamantes [0.50%], Diapherotrites [0.39%], Acetothermia [0.37%], Synergistetes [0.24%], Gracilibacteria [0.22%], Caldiserica [0.22%], Euryarchaeota [0.21%], Aenigmarchaeota [0.17%], Thermotogae [0.09%], Thaumarchaeota [0.08%], WS1 [0.05%]). Secondly, the origins of these reads, i.e. their assignment before the addition of the SAG dataset, were determined (Bacteria [36.06%], cellular-organisms [23.42%], No-hits [10.22%], Proteobacteria [5.64%], Bacteroidetes/Chlorobi-group [5.24%], unclassified-Bacteria [2.92%], Firmicutes [2.89%], Opisthokonta [1.88%], root [1.37%], Euryarchaeota [1.34%], Eukaryota [0.89%], Archaea [0.87%], prokaryotic-environmental-samples [0.79%], Fibrobacteres/Acidobacteria-group [0.75%], Planctomycetes [0.74%], Actinobacteria-phylum [0.59%], Chloroflexi-phylum [0.44%], environmental-samples-Bacteria [0.43%], Cyanobacteria [0.42%], Chlamydiae/Verrucomicrobia-group [0.31%], Viridiplantae [0.28%], Spirochaetes [0.26%], Thaumarchaeota [0.21%], Crenarchaeota [0.19%], environmental-samples-Archaea [0.17%], Alveolata [0.15%], Aquificae-phylum [0.13%], Deinococcus-Thermus [0.10%], Thermotogae-phylum [0.09%], environmental-samples-unclassified [0.05%], Amoebozoa [0.03%], stramenopiles [0.02%], Euglenozoa [0.02%], Nitrospirae [0.02%], Synergistetes [0.02%], dsDNA-viruses-no-RNA-stage [0.02%], Deferribacteres-phylum [0.02%], Not-assigned [0.02%], Fusobacteria [0.02%], Viruses [0.01%], Dictyoglomi [0.01%], Tenericutes [0.01%], Thermodesulfobacteria-phylum [0.01%], and Caudovirales [0.01%]).

Phylogenetic diversity

The phylogenetic diversity (PD; **Supplementary Fig. 20**) values were calculated from a SSU rRNA maximum likelihood tree (fasttree, GTR model, CAT approximation) based on the 99% non-redundant greengenes database (release 10-2012). The PD was determined by successively traversing the tree for each group in the order 'Non-GEBA genomes', 'GEBA genomes', and 'Microbial Dark Matter genomes'. The tree was then further traversed for both of the groups 'Organisms from the greengenes database excluding environmental samples' and 'Organisms from the greengenes database' in parallel and the results were later overlaid in the final graphic. For each leaf in its current group, the branch lengths of previously unvisited paths leading from the root node to the leaf was summed up, representing the PD of this leaf. Then this PD value was added to the sum of all previously traversed leafs and plotted against the number of all visited leafs. The calculations have been performed by a custom perl program.

Novel protein families

Novel protein families (**Supplementary Fig. 21**) were determined by performing all against all BLASTP searches of the protein sequences (e-value cutoff of $1e^{-3}$ was applied) before processing the results with the Markov cluster algorithm (Enright et al., 2002) using default settings. Each resulting cluster of proteins was searched against the Pfam database (Finn et al., 2009) using HMMER3 software (Eddy, 2011) and considered 'novel' when more than 50% of the containing proteins did not result in any hit (trusted cutoff of Pfam HMM models was applied). For the creation of the rarefaction curve, this procedure was repeated 10 times for every step starting with 1000 proteins and continuously adding 1000 more randomly selected proteins until the final step in which all 94,601 proteins (maximum of the MDM dataset) were processed at once. Mean values and standard deviations were calculated for each step. In the case of the GEBA and IMG isolate datasets, 94,601 randomly selected proteins were chosen from the pools of 100 randomly selected genomes.

Gene neighborhood analysis of genes involved in novel findings

For the detailed assembly plots, corresponding sequence regions were extracted from the final assembly fasta files. Using the BWA (Li and Durbin, 2009) reference alignment software, all sequenced reads were mapped against the extracted region and saved as files in the SAM format (Li et al., 2009). Custom perl scripts were used to parse the SAM file and generate the visualizations (**Supplementary Fig. 25**).

References

- Darling A, H Bik, E Lowe, G Jospin, J A Eisen. PhyloSift: metagenome phylogeny for the 1%. In preparation.
- Dean, F. B., Nelson, J. R., Giesler, T. L. & Lasken, R. S. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11**, 1095–1099 (2001).
- Eddy S. HMMER 3.0 (March 2010); <http://hmmer.org/>
- Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol.* 2011 Oct;7(10):e1002195
- Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
- Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Research* **38**, D211–D222 (2009).
- Gnerre S, MacCallum I. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *PNAS* 2011;108:4 1513–1518.
- Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Research* **21**, 1552–1560 (2011).
- Hyatt D, Chen GL, Lacascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; 11:119.
- Katoh, K. & Frith, M. C. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**, 3144–3146 (2012).
- Kielbasa SM, R Wan, K Sato, P Horton, MC Frith. Adaptive seeds tame genomic sequence comparison. *Genome Research* 2011.
- Kunin V. and P. Hugenholtz. PyroTagger: A fast, accurate pipeline for analysis of rRNA amplicon pyrosequence data. *The Open Journal*, 1, Feb 2010.
- Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* **39**, W475–W478 (2011).
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964.
- Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Research* **32**, 1363 – 1371 (2004).
- Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* 2013 Jan 1;41(D1):D348-52.

- Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; 25:2271-2278
- Markowitz, Victor M., I-Min A. Chen, Krishna Palaniappan, Ken Chu, Ernest Szeto, Yuri Grechkin, Anna Ratner, et al. "IMG: The Integrated Microbial Genomes Database and Comparative Analysis System." *Nucleic Acids Research* 40, no. D1 (January 2012): D115–D122.
- McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 6, 610–618 (2012).
- Owen, M. & Provan, J. S. A Fast Algorithm for Computing Geodesic Distances in Tree Space. *arXiv:0907.3942* (2009). at <<http://arxiv.org/abs/0907.3942>>
- Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; 7:455-457.
- Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* 5, e9490 (2010).
- Pruesse E, Quast C, Knittel, Fuchs B, Ludwig W, Peplies J, Glöckner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nuc Acids Res* 2007; 35: 2188-7196.
- Punta, M., P.C. Coghill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. Finn The Pfam protein families database. *Nucleic Acids Research* (2012) Database Issue 40:D290-D301
- Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences* (2009). doi:10.1073/pnas.0906412106
- Stamatakis, A. *et al.* RAxML-Light: A Tool for computing TeraByte Phylogenies. *Bioinformatics* (2012).doi:10.1093/bioinformatics/bts309
- Swan, B. K. *et al.* Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* 333, 1296 –1300 (2011).
- White, R. A., Blainey, P. C., Fan, H. C. & Quake, S. R. Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* 10, 116 (2009).
- Woyke, T. *et al.* Decontamination of MDA Reagents for Single Cell Whole Genome Amplification. *PLoS ONE* 6, e26161 (2011).
- Wu D, G Jospin, J A Eisen. Systematic identification of phylogenetic markers at different taxonomic levels for bacteria and archaea. In preparation.
- Zerbino D, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008;18:821–829.

Supplementary Text

Genome-based phylogenetic inference

FCB superphylum

Comparative analysis of over 5,000 genomes within IMG revealed that a conserved C-terminal domain (CTD) of extracellular proteinases (TIGR04183; Por secretion system C-terminal sorting domain) is exclusively found in members of the FCB superphylum. This includes the original members Fibrobacteres, Chlorobi, Bacteriodes, as well as the candidate phyla Cloacimonetes (WWE1), Marinimicrobia (SAR406), Latescibacteria (WS3) and the only available genome of the genus *Caldithrix*. The conserved C-terminal domain of interest is however not found in the only Gemmatimonadetes genome available (*Gemmatimonas aurantiaca* T-27T) and occurs in only one of the 12 sequenced genomes of the phylum Chlorobium. Therefore we suggest that the presence of this CTD is an exclusive but not obligate feature of the FCB superphylum. The only significant hits (TIGR trusted cutoff=23) outside of the FCB superphylum were to Cyanobacteria (genus *Arthrospira*, *Trichodesmium*, *Synechococcus*) to domains located in the middle of genes, which are therefore assumed to be nonfunctional.

PVC superphylum

The PVC superphylum named after the phyla Planctomycetes-Verrucomicrobia-Chlamydiae (PVC) based on SSU rRNA analysis, was proposed to also include the phylum Omnitrophica (OP3) and the phylum Lentisphaerae (Wagner and Horn, 2006). A recently described conserved signature gene of unknown function was exclusively found in all sequenced species belonging to the phyla Chlamydiae, Verrucomicrobia, Lentisphaerae, and Planctomycetes (Gupta et al., 2012). The gene was originally described as protein CT421.2 from *C. trachomatis* (accession number NP_219933) and it was subsequently found that the length varies from 53 aa in the Chlamydiaceae to more than 80 aa in the Planctomycetes. We could detect this gene encoding a hypothetical protein (TIGR04137) in one of our four Omnitrophica genomes (Locus Tag: A11A17DRAFT_01756) with a length of 83 aa.

DPANN superphylum

The proposed phyla Diapherotrites (pMC2A384), Parvarchaeota, Aenigmarchaeota (DSEG) and Nanoarchaeota together with the Nanoarchaeota form a monophyletic superphylum in Archaea-only trees for which we propose the identifier, DPANN. We predict that small cell and genome size are unifying features of the DPANN phyla. However, when adding bacterial outgroups several lineages of the superphylum DPANN are pulled towards Bacteria (**Supplementary Table 4b**) which could be attributed to Long branch attraction (LBA) of DPANN lineages to the bacterial outgroup. Thus we emphasize that the proposed superphylum DPANN needs to be further evaluated as more genomes and improved phylogenetic inference methods come to hand.

Justification of phylogenetic inference methods

Concatenated gene alignments and alternatives to calculate the TOL

The question if superalignment (concatenating the gene alignments; also called "supermatrix", "concatenation" or "total evidence") or supertree (combining the data at the level of gene trees) approaches are more accurate to reconstruct phylogenies has been discussed extensively in the literature. For example, Gadagkar et al. (2005) concluded that the concatenation approach yields more accurate trees, even when the sequences concatenated have evolved with very different substitution patterns. Kupczok et al. (2010) found that a superalignment usually performs best since it is less susceptible to stochastic error but that supertree methods can outperform superalignments in the presence of gene-tree conflicts. An extensive phylogenetic inference including over 3000 microbial (bacterial and archaeal) genomes constructed with supertrees and supermatrixes (Morgan-Lang et al., 2013) concludes that the best approach for generating a single phylogenetic tree, suitable for use as a reference phylogeny for comparative analyses, is to perform a Maximum Likelihood (ML) analysis of a concatenated alignment of conserved, single-copy genes. Indubitably, the selection of gene sequences used to build the concatenated alignment influences the accuracy of the phylogenetic inference. All of our 38 markers are universally distributed and exist as single copy genes, 35 of them have previously been tested and were found to be relatively recalcitrant to lateral gene transfers (Wu and Eisen, 2008; Wu and Scott, 2012) and 31 were successfully used for large scale phylogenetic inferences (Wu et al., 2009, *Nature*). The majority of our markers (24 genes) was also extensively tested by Morgan-Lang et al. 2013 and found to be most suitable for phylogenetic analysis. However to evaluate if the individual genes have similar phylogenies, we calculated individual gene phylogenies for all 38 genes and used distributions of pairwise geodesic distances between those gene trees to assess levels of congruence (Supplementary Methods). The resulting frequency plot shows no outliers, which suggests that all individual gene trees represent comparable branch lengths and topologies, and the hierarchical clustering of pairwise geodesic distances shows a congruent clustering pattern resulting in an additive phylogenetic signal of the concatenated gene tree (**Supplementary Fig. 7**).

Long branch attraction

Long branch attraction (LBA) is a well-known phylogenetic artifact that causes erroneous grouping of two or more long branches (Bergsten, 2005). This artifact was originally thought to mainly affect maximum parsimony analyses, but recent simulation studies have shown that even Bayesian inference can be biased in favor of topologies that group long branches together (Kolaczkowski and Thornton, 2009). We applied ML methods, which have various implemented measures to avoid LBA artifacts. FastTree2 includes SPR (subtree-pruning-regrafting) moves with log-corrected distances to improve the starting tree, which should correct any LBA that occurred during the Neighbor Joining (NJ) phase (Price et al., 2010). FastTree2 also performs maximum likelihood Nearest Neighbor Interchanges (NNIs) aiming to further eliminate LBA (Price et al., 2010). RAxML is thought to be even less susceptible to LBA as a consequence of the search algorithm being used because it conducts a more thorough and thus exhaustive ML-based optimization of the tree topology (Alexandros Stamatakis pers. comm.). However no model can be assumed to be entirely correct for real data, and LBAs cannot be excluded for any phylogenetic inference method. Therefore one should implement measures to reduce LBA which include sampling more taxa to break up long branches, using inference methods less sensitive to LBA (such as likelihood), and sampling more characters, both in number and variety (Bergsten, 2005). We used ML inference methods (see above) and built our phylogeny with sequence alignments of over 2000 taxa, in order to reduce LBA. Furthermore our calculations are based on amino acid sequences since the possibility of obtaining LBA decrease with an increasing alphabet of character states. Another common problem is LBA to outgroups. It is argued that since outgroup taxa

almost always represent long branches and are as such a hazard towards attracting long branched ingroup taxa, phylogenetic analyses should always be run with and without outgroups included (Bergsten, 2005). Such a measure will show if the outgroup roots the ingroup or if it simultaneously alters the ingroup topology. Therefore we performed phylogenetic analysis of the individual domains to assess inter-domain LBA (**Supplementary Methods, Supplementary Table 4**).

Large datasets

Genome comparison and modern inferences of deep phylogenies using conserved proteins rely almost exclusively on ML and Bayesian methods (Yang and Rannala, 2012). Both methods share many statistical properties such as consistency and efficiency, and they have a variety of sophisticated evolutionary models available. A major challenge of both methods is computation. Recent developments of ML tools for large-scale phylogenetic inference on supercomputers aimed to solve this gridlock. For example FastTree2 was shown to be 100–1,000 times faster than traditional likelihood methods, however at the cost that it is presumably not quite as accurate as methods that use maximum-likelihood SPRs (Price et al., 2010). However, a recent study shows that very large phylogenies can be estimated very quickly using FastTree, with little (and in some cases no) degradation in tree accuracy, as compared to RAxML (Liu et al., 2011). Also RAxML-light (Stamatakis et al., 2012) demonstrated scalability and robustness when run on multiple computer nodes with a simulated DNA alignment of 1,481 taxa and 20,000,000 bp. The popular Bayesian inference software Mr. Bayes however is not as easily scalable as tested by a recent study which included 841 organisms and 24 marker genes and found that, after approximately 18 months of CPU time and two million MCMC generations the analysis did not yet converge upon an answer (Morgan-Lang et al., 2013). Therefore for our 8,102 character alignments of more than 2,000 taxa we found Fasttree2 and RAxML to be the most suitable inference methods.

Incomplete datasets

Many genomic data sets, including a subset of our own (**Supplementary Table 2**), are incomplete, and although in theory ML and Bayesian methods can accommodate missing data, the impact of large-scale missing data is not well understood (Yang and Rannala, 2012). However several studies have implicitly made the assumption that a certain amount of missing data (up to 25%) can be tolerated by tree-reconstruction methods, and simulations as well as empirical approaches have shown that this proportion can be surprisingly high without losing substantial accuracy (Delscu et al., 2005). It is understood that the impact of missing data is limited because species for which sequence information is incomplete are still represented by a large number of informative characters in phylogenomic inferences (Philippe et al., 2004). When applying ML methods Philippe et al. found that the inferred phylogeny remained similar to the model, even when 90% of the data were missing. Thus, it has been suggested that having a species represented by only 10 genes out of 100 that are analyzed in the entire study might generally be less problematic than not considering that species at all (Delscu et al., 2005). In order to test for a possible bias introduced by short sequences, we calculated trees including only taxa with certain alignment lengths and marker genes, starting from full alignment lengths and adding short sequences in two steps (**Supplementary Methods**). Adding short sequences did not compromise the monophyly of candidate phyla (**Supplementary Table 4**).

Functional Diversity and Novel Findings

Our single cell genomes allowed us to provide a first glimpse into the potential functional diversity of many of the investigated candidate phyla and novel lineages. Hydrogen metabolism is widespread amongst the novel lineages which is evident by hydrogenase encoding genes. Hydrogenase proteins catalyze the reversible conversion of molecular hydrogen to protons and electrons. The most abundant hydrogenases contain a [NiFe] active site and are generally biased towards hydrogen oxidation activity and are reversibly inhibited by oxygen, however exceptions exist (Shafaat et al., 2013). Enzymes of the group 1 NiFe hydrogenases detected in BRC1 and Aerophobetes (CD12) are membrane-bound respiratory enzymes that channel electrons from hydrogen into the quinone pool, providing the link between hydrogen oxidation and energy production (Supplementary Table 4).

Our analysis indicates that Aerophobetes (CD12), Omnitrophica (OP3) SAGs, and one Cloacimonetes (WWE1) SAG are capable of fixing carbon via the Reductive acetyl-CoA cycle (Wood-Ljungdahl pathway; **Supplementary Fig. 12, Supplementary Table 4**). Putative Ribulose-1,5-bisphosphate Carboxylase Oxygenase (RuBisCO) genes were identified in several SAGs (**Supplementary Fig. 12; Supplementary Table 4**). Five genes, from archaeal SAGs and the bacterial phylum Parcubacteria (OD1) could be assigned to RuBisCO type III (**Supplementary Fig. 22**) which does not function in the Calvin-Benson-Cycle but fixes CO₂ via the adenosine nucleotide degradation pathway. Type III RuBisCO was originally described in Archaea but has recently been found in bacterial genomes as well (Wrighton et al., 2012). The putative RuBisCO gene of the Aigarchaeum SAG (0000106-J15) clusters within type IV (the RuBisCO like protein, **Supplementary Fig. 22**).

The great phylogenetic distance of our single-cell genomes to previously sequenced microbes allowed us to find several surprising discoveries including a reassignment of codon usage, sigma factors in Archaea, horizontal gene transfer across domains and evidence for archaeal pathogenesis.

Codon reassignment

The G+C content of bacterial genomes ranges from 16% to 75% and appears to be the primary determinant of the codon and amino acid usage patterns observed in bacterial genomes (Lightfield et al., 2011). Thus microbes with a high genomic G+C content above 70% almost exclusively use G- or C-ending codons, whereas low G+C organisms with a G+C below 30% heavily use A- and U-ending codons when possible (Lafay et al., 2000). This poses a challenge for organisms with extreme G+C content, and it has been suggested that a high magnitude reduction in the genomic G+C content would require a severe reduction in the use of amino acids such as proline, alanine, and glycine that are encoded by GC-rich codons (Lightfield et al., 2011; **Supplementary Fig. 13c**). We discovered that two SAGs from phylum Gracilibacteria (GN02, BD1-5) found a different solution for this challenge - by altering their codon usage pattern. The codon UGA (amber) is not used as a nonsense or termination codon but as a codon for the amino acid glycine. This conclusion is based on the evidence that two Gracilibacteria genomes include UGA codons in their reading frames and they appear at positions corresponding to glycine in homologous genes (**Supplementary Fig. 13a**). Such recoding event has very recently also been found in the candidate phylum SR1 from the oral mouth cavity (Campbell et al., 2013). Campbell et al. also detected UGA codon reassignment to glycine in the SR1-related ACD80 and other members of the Gracilibacteria (ACD78 within BD1-5/GN02) and biochemically validated these sequence-based findings. Other alternative codon

usage patterns involving UGA are known. A variety of organisms from all three domains of life have the amino acid selenocysteine encoded by UGA (Johansson et al., 2005). In addition to this more common alternative usage, some eukaryotes have mitochondria which use UGA to encode tryptophan (Trp) rather than a chain terminator (Sengupta et al., 2007). The only examples from the bacterial world are the pathogen *Mycoplasma capricolum* (Mollicutes; Yamao et al., 1985), and the alphaproteobacterial symbiont of cicadas (McCutcheon et al., 2009) which also use both the UGG and UGA codons for tryptophan. Besides the exceptional symbiotic alphaproteobacterium, which has an extremely small genome (144 kb) and a higher G+C content (McCutcheon et al., 2009), it is believed that the base composition is the key driver for the codon reassignment. Both the *M. capricolum* and two Gracilibacteria genomes are characterized by an extremely low G+C content (25, 22 and 23%).

The reasons for an exceptional low G+C content as well as for the variation among genomic G+C contents in general remain controversial and recent studies suggest an universal mutation bias towards AT, even in species which have high genomic G+C content (Hershberg and Petrov, 2010; Hildebrand et al., 2010) and present evidence that the DNA polymerase III α -subunit and its isoforms play a dominant role in determining the G+C content (Wu et al., 2012). An analysis of 2 Gbp of metagenome data from several sampling sites showed that bacterial communities from particular environments have surprisingly narrow G+C content distributions and that these distributions differ between environments, independent of distinct phylogenetic origins of the species (Foerstner et al., 2005). This suggests that certain G+C contents are favored in particular environments. In the case of the Gracilibacteria genomes, which were sampled in 2,503 m depth at the hydrothermal vents of the East Pacific rise (EPR), there could be several possibilities: The first is an adaptation to a hyperthermophilic lifestyle: A comparison of the genomic G+C content versus growth temperatures shows a trend of increasing G+C with increasing temperature preference for most organisms, except for hyperthermophiles - organism thriving in extremely hot environments from 60 °C upwards (**Supplementary Fig. 23**). The turbulent mixing conditions and steep gradients at hydrothermal vents make it challenging to predict if the Gracilibacteria SAGs live around the vents at 25 °C temperatures where they were sampled, or are thermophiles originating from the seafloor or hydrothermal fluids. We did not find genes encoding reverse gyrase (COG1110), an enzyme preventing the DNA from supercoiling at high temperatures which is closely tied to hyperthermophilic life (Heine and Chandra, 2009), in any of the Gracilibacteria SAGs. Applying the well-established correlation between the G+C content of the paired stem regions of the 16S rRNA genes and the growth temperature (Wang et al., 2006) suggest that the Gracilibacteria SAGs are mesophiles and that their optimal growth temperature is at or below the measured 25 °C at the sampling site (**Supplementary Fig. 24**).

Another possibility for the low G+C content of the two Gracilibacteria SAGs could be the limited availability of nitrogen. Since A/T bases consist of less nitrogen than G/C bases such nitrogen limitation would favor a low G+C content. In addition Glycine is amongst the amino acids which require the lowest number of nitrogen atoms (**Supplementary Fig. 13d**). Dissolved dinitrogen gas (N₂), is abundant in deep seawater (0.59 mM), slightly elevated in hydrothermal fluids (Charlou et al., 2000) and recent studies showed nitrogen fixation is possible at up to 92 °C. However we did not find any genes coding for nitrogen fixing enzymes in the Gracilibacteria SAGs. While we did not recover the entire genomes, the lack of *nif* genes is nonetheless suggestive of the inability of Gracilibacteria to utilize N₂, which could be the potential driver for its low genomic G+C content.

The standard IMG annotation did not detect any UGA specific t-RNAs in the two Gracilibacteria genomes; however an RNA analysis for matches against the Rfam database (Burge et al., 2012)

successfully recovered a tRNA in SAG 277 (K10), which has an anti-codon of UGA. It is encoded at the end of contig 20_0.21 (coordinates: 8335-8260) and has a good bit-score (44.1; e-value 2.3e^{-10}). A second hit with a lower bit score (26.8; e-value 7.5e^{-06}) to selenocysteine was also recovered, but a detailed analysis of the alignment showed that we can rule out the possibility of it being a tRNA for selenocysteine, since the major identity determinant for selenocysteine biosynthesis and insertion, the 13 bp long extended acceptor arm is missing (Commans and Böck, 1999). We did not find a release factor RF2 which recognizes UGA in any of the two recoded *Gracilibacteria* genomes.

Purine synthesis

We found another unusual gene composition with high similarities to archaeal genes among three SAGs belonging to the Microgenomates (OP11). The synthesis of purines, which are key components of all living cells for energy metabolism (ATP) and the bio-synthesis of RNA and DNA, is quite similar and well conserved in Bacteria and Eukaryotes whereas the archaeal pathway is more variable (Brown et al., 2011). A main difference between the bacterial and archaeal pathway is that the majority of Archaea utilizes the PurP protein, whereas Bacteria use the proteins PurH1 to formylate the ribonucleate. None of the current bacterial genomes in IMG encodes the PurP protein. Surprisingly we found a nearly complete purine synthesis operon in three Microgenomates SAGs (**Supplementary Fig. 14**) and all three genomes encode the PurP protein, which could be due to a lateral gene transfer from an euryarchaeal donor. The purin synthesis pathway in our Microgenomates SAGs is almost complete, missing only the last step to cyclize and to reform the purine ring, a function which is carried out by PurO in Archaea and PurH2 in Bacteria. Since we did not find any homologs to PurO or PurH2 in any of the Microgenomates SAGs there could possibly be a third unrecognized version of the enzyme, or alternatively we missed this part of the genome.

Lateral gene transfer across domains

Horizontal gene transfer (HGT) refers to the transfer of genetic material between organisms other than from the parental generation to the offspring. There is a growing awareness that horizontal gene transfer is a highly significant phenomenon and amongst microbes perhaps the dominant form of genetic transfer, as shown for marine bacteria (Mc Daniel et al., 2001). HGT across domains has been demonstrated (see Boto, 2010; even involving eukaryotes). For example the bacterial genus *Agrobacterium* uses a plasmid to transfer DNA to plant somatic cells causing large woody protrusions, and *Wolbachia* an alphaproteobacterial endosymbiont has transferred 11 kb of DNA to the X chromosome of its insect host (see Brown, 2003). It is believed that HGT was extensive in the early stages of cellular evolution, when the barriers to genetic exchange between the Archaea, the Bacteria and the eukaryotes were weak (Brown, 2003), however it is rare to have an eukaryotic protein transferred to Bacteria (Andersson, 2005). A recent example was found from an insect to a bacterial endosymbiont (Woolfit et al., 2009). To our knowledge no HGT from Eukaryotes to Archaea has been documented so far.

Here we report a HGT event between an ancient Archaeon and an Eukaryote. A probable oxidoreductase from a Nanoarchaeota SAG is most closely related to an orthologue of the soil amoeba *Dictyostelium discoideum* and related slime molds (**Supplementary Fig. 15**). HGT from “prokaryotes” to slime molds has been reported. Analysis of the complete genome sequence of *D. discoideum* showed that HGT is a potentially important mechanism of its genome evolution, and led to the identification of 18 genes derived from Bacteria (Eichinger et al., 2005). HGT from Eukaryotes to Archaea seems to be uncommon, since it has been suggested as a possibility only for a putative zinc-chelating domain based on

an analysis of 18 genomes of all three domains of life (Ponting et al., 1999). However a search for IPR000058 (Zinc finger, AN1-type) against all genomes in IMG shows the domain is present in 225 genomes (55 Archaea, 3 Bacteria, 162 Eukaryotes, 5 Viruses) which may invoke the conclusion that this domain is most likely a widely used component of proteins in all three domains of life and not a recent HGT. Thus to our knowledge no phylogenetic evidence for a HGT of a complete gene from Eukaryotes to Archaea has been reported so far.

Sigma factors

Transcription is the first step of gene expression and is accomplished by multi-subunit RNA polymerases in all living organisms. The initiation of transcription in Bacteria is carried out by RNA polymerase in association with a sigma factor. The sigma factor recognizes the promoter DNA sequence and alters the structure of the RNA polymerase to promote initiation (reviewed by Gross et al., 1998). Most bacteria contain multiple sigma factors that enable the recognition of different sets of promoters. One covers the housekeeping functions and is generally known as the major or primary sigma factor, whereas others are used for more specialized functions. With regard to sequence similarity, sigma factors can be grouped into two classes, the sigma-54 and sigma-70 families. Sigma factors of the sigma-70 family vary in size and domain composition. For example the *E. coli* housekeeping sigma factor contains four regions, region 1 to 4, while the smaller extracytoplasmic function (ECF) sigma factors contain only regions 2 and 4. Region 2 recognizes the -10 region of the promoter and region 4 recognizes the -35 region (**Fig. 3**). Contacts between the sigma factor and RNA polymerase involve regions 2, 3, and 4 (reviewed by Gruber and Gross, 2003).

Archaea lack sigma factors and use a simplified version of the eukaryotic system for transcription initiation. The three major factors involved in archaeal transcription initiation are the TATA-binding protein (TBP), which binds the TATA box of the promoter, transcription factor B (TFB) which interacts with another region of the promoter located immediately upstream of the T/A box, and transcription factor E (TFE). The archaeal RNA polymerase is more similar to the eukaryotic enzyme, but also shares sequence similarity with bacterial RNA polymerases (Thomm, 2007). Numerous archaea have proteins with the sigma factor region 4 domain, a fact first noted by Kyrpides and Ouzounis (1997). Most of the archaeal proteins with this domain belong to the Tfx family of DNA binding proteins (TIGR00721) specific to the archaea. The Tfx protein from *Methanobacterium thermoautotrophicum* has properties consistent with a role as a transcription factor (Hochheimer et al., 1999) [†].

Mining the archaeal SAGs, we discovered numerous proteins with sigma factor domains, some of which have similarity over the entire length of the protein to bacterial sigma factors of the sigma-70 family. Thus our findings challenge the current view that sigma factors are not used for archaeal transcription. Phylogenetic analysis of the conserved regions 2 and 4 (**Fig. 3, Supplementary Fig. 16**) shows that a protein from one of the Diapherotrites (pMC2A384) genomes belongs to group 3, whereas another Diapherotrites protein branches close to the base of Group 4 sigma factors. Three genes from Nanoarchaeota clearly belong to Group 4. The genomes that contain sigma factors also contain at least one of the three archaeal basal transcription factors (TBP, TFB, TFE), except in the case of Nanoarchaeota, where two of the genomes with the Group 4 sigma factor lack the archaeal transcription factors while the other one contains all three.

In the phylum Nanoarchaeota, three single-cell genomes contain a protein that can be assigned to ECF family ECF01, which includes the *B. subtilis* protein sigma(W) (using the web server ECF finder,

<http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder/>). Many ECF sigma factors are maintained in an inactive state by an anti-sigma factor until a signal is perceived by the cell. The anti-sigma factor is often encoded in an operon with the sigma factor and located downstream from the sigma factor. Anti-sigma factors are diverse in sequence: some contain a transmembrane domain and are activated by an external signal, while some are cytoplasmic, and some ECF sigma factors do not appear to have anti-sigma factors (Staroń et al., 2009). The Nanoarchaeota SAGs with the ECF01 family sigma factor did not have genes with similarity to any of the known ECF01 anti-sigma factors, and the genes downstream of the sigma factors do not have significant similarity to any other known protein, so we cannot determine if an anti-sigma factor is present in these genomes from sequence analysis alone.

We can only speculate about the function of the sigma factors within archaeal genomes. It is possible that they are involved in transcription initiation for a subset of genes and/or in response to different conditions. *E. coli* for example has, next to the main sigma factor, six other sigma factors that accumulate in response to specific stresses and allow to initiate transcription at specific promoters (Ishihama, 2000). Especially sigma factors of the extracytoplasmic function (ECF) family, which we found in the archaeal lineages (Nanoarchaeota and Diapherotrites), have been shown to regulate gene expression in response to sensing a variety of signals from outside the cell (Brooks and Buchanan, 2008). Alternatively the archaeal sigma factors could work as DNA-binding proteins with a different function, such as transcription factors, and in such a case they would not replace the archaeal basal transcription factors' role in transcription initiation.

Stringent response

When bacteria face nutrient limitations or stress in their environment threatens their survival they have to adapt to these conditions quickly, for example with the stringent response. During the stringent response bacteria can economise nutrients by lowering their metabolism to a minimum until the environment improves (Wu and Xie, 2009). The signaling molecules pppGpp and ppGpp (Guanosine 5'-(tri)diphosphate, 3'-diphosphate), known as alarmones, mediate the stringent response in bacteria and affect numerous cell properties, with different effects in different organisms (Dalebroux and Swanson, 2012). Alarmone synthesis involves generation of pppGpp by a synthase, followed by removal of one phosphate to form ppGpp. *E. coli* contains two pppGpp synthases: RelA is involved in the stringent response to amino acid limitation, while SpoT responds to other stress conditions (Dalebroux and Swanson, 2012). RelA and SpoT are multidomain proteins containing a (p)ppGpp hydrolase domain (nonfunctional in RelA) and a pppGpp synthase domain in the N terminus, with a TGS domain and an ACT domain in the C terminus. Most bacteria have a single homolog of RelA and SpoT. The C terminus is involved in sensing stress conditions and regulating the pppGpp synthase and (p)ppGpp hydrolase activities (Avarbock et al., 2005; Battesti and Bouveret, 2006). Recently smaller, single-domain pppGpp synthases have been identified in some bacteria; since they lack a regulatory domain, they may be regulated at the level of transcription (Tozawa and Nomura, 2011). Production of ppGpp also occurs in chloroplasts of plants and algae, and homologs of RelA/SpoT are found in these organisms (Tozawa and Nomura, 2011). Recently a protein with ppGpp hydrolase activity was identified in human and *Drosophila*, but ppGpp itself has not been detected in animals (Sun et al., 2010).

A few archaea have been found to have a stringent response; however, ppGpp was not found in any of these Archaea (Cellini et al., 2004). Recently an extensive sequence analysis showed that a putative single domain ppGpp synthase is present in *Methanosarcina acetivorans*, while putative single domain ppGpp hydrolases were found in four euryarchaeotes; however, no archaeon was found that had both a synthase and a hydrolase (Atkinson et al., 2011). Two of the single-cell archaeal genomes presented here, one belonging to the phylum Diapherotrites (pMC2A384) and one to the Nanoarchaeota, encode full length homologs of RelA/SpoT having a synthase domain, a hydrolase domain, and a regulatory domain, providing strong evidence that some archaea can produce ppGpp in response to the sensation of an intracellular signal (**Supplementary Fig. 17**). The Diapherotrites genome contains two RelA/SpoT homologs, suggesting that more than one internal signal can control ppGpp synthesis in this organism. Since Diapherotrites and Nanoarchaeota branch deeply within the Archaea, we hypothesize that the earliest Archaea, like most Bacteria, were capable of ppGpp synthesis as a response to nutrient deprivation, and that this capacity was lost early within archaeal evolution. This may have occurred as the Archaea were out-competed by Bacteria in most environments and became specialized to extreme or nutrient poor environments or developed unique metabolic pathways such as methanogenesis. In these niches competition is reduced, but energy stress is common (Valentine, 2007). As a result, many Archaea grow slowly, and many also have only one ribosomal RNA operon, which correlates with a slow response to nutrient availability (Klappenbach et al., 2000), so the need for a global response to nutrient deprivation may have become unnecessary.

Archaeal toxins - lytic murein transglycosylases

The questions if pathogenic Archaea truly exist and why none have been identified so far remains a hot topic in biomedical research (Gill and Brinkman, 2011). Pathogenesis is in many diseases a complex

process, however some only require the secretion of a single toxin as with the bacterium *Clostridium botulinum* (Cai et al., 2007). Archaea have a range of toxins at their disposal but these archaeocins tend to target other Archaea and have not been shown to inhibit any bacterial organism (Shand and Leyva, 2007). We hypothesize for the first time that Archaea may produce compounds which are effective against members of the domain Bacteria.

We discovered a gene encoding a lytic murein transglycosylases in two archaeal SAGs (Nanoarchaeota SAGs AAA011-L22 and AAA011-K22; **Supplementary Fig. 18**). Both have a complete catalytic domain, and are predicted to have signal peptides allowing proteins to be secreted through the universally conserved Sec pathway. Murein transglycosylases degrade murein (peptidoglycan) strands, which are the main components of the bacterial cell wall, by cleaving the glycosidic linkage between *N*-acetylmuramoyl and *N*-acetylglucosaminyl residues (Recolons et al., 2011). They are ubiquitous in bacteria and responsible for creating space within the peptidoglycan sacculus for its biosynthesis, recycling and cell division and have been suggested to contribute to pathogenesis (Scheurwater et al., 2008).

[†] **note:** In addition we found a RNA polymerase sigma factor homolog annotated on one of the 171 contigs of the “Candidatus Nitrosopumilus salaria BD31” draft genome (UniProtKB I3D119). Interestingly the best blast hits of this and other genes on the same contig are to the betaproteobacterial genus *Methylobacter*. The contig in question also contains bacterial genes such as an outer membrane protein which is a signature of gram-negative bacteria (such as Betaproteobacteria), and the cell division ATP-binding protein FtsE which is unique to Bacteria. We performed tetranucleotide analysis of the “Candidatus Nitrosopumilus salaria BD31” draft genome (data not shown) and found that several contigs including the one with the RNA polymerase sigma factor homolog were far removed from the central cloud which further supports that these contigs are not part of the genome and are a possible contamination.

References

- Andersson, J. O. Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences* **62**, 1182–1197 (2005).
- Atkinson GC, Tenson T, Hauryliuk V. The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. *PLoS One* 2011;6:e23479.
- Avarbock A, Avarbock D, Teh JS, Buckstein M, Wang ZM, Rubin H. Functional regulation of the opposing (p)ppGpp synthetase/hydrolase activities of RelMtb from *Mycobacterium tuberculosis*. *Biochemistry* 2005;44:9913-9923.
- Battesti A, Bouveret E. Acyl carrier protein/SpoT interaction, the switch linking SpoT-dependent stress response to fatty acid metabolism. *Mol Microbiol* 2006;62:1048-1063.
- Bergsten, J. A review of long-branch attraction. *Cladistics* 21, 163–193 (2005).
- Boto, L. Horizontal gene transfer in evolution: facts and challenges. *Proc. R. Soc. B* **277**, 819–827 (2010).
- Brooks, B. E. & Buchanan, S. K. Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1778**, 1930–1945 (2008).

- Brown, J. R. Ancient horizontal gene transfer. *Nature Reviews Genetics* **4**, 121–132 (2003).
- Burge, S. W. *et al.* Rfam 11.0: 10 years of RNA families. *Nucl. Acids Res.* (2012). doi:10.1093/nar/gks1005
- Cai S, Singh BR, Sharma S. Botulism diagnostics: from clinical symptoms to in vitro assays. *Crit Rev Microbiol.* (2007) Apr-Jun;33(2):109-25. Review.
- Campbell, J. H. *et al.* UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *PNAS* **110**, 5540–5545 (2013).
- Cavicchioli, R., Curmi, P. M. G., Saunders, N. & Thomas, T. Pathogenic archaea: do they exist? *BioEssays* **25**, 1119–1128 (2003).
- Cellini A, Scoarughi GL, Poggiali P, Santino I, Sessa R, Donini P, Cimmino C. Stringent control in the archaeal genus *Sulfolobus*. *Res Microbiol* 2004;155:98-104.
- Charlou, J. . *et al.* Compared geochemical signatures and the evolution of Menez Gwen (37°50'N) and Lucky Strike (37°17'N) hydrothermal fluids, south of the Azores Triple Junction on the Mid-Atlantic Ridge. *Chemical Geology* **171**, 49–75 (2000).
- Commans, S. & Böck, A. Selenocysteine inserting tRNAs: an overview. *FEMS Microbiol. Rev.* **23**, 335–351 (1999).
- Dalebroux ZD, Swanson MS. ppGpp: magic beyond RNA polymerase. *Nat Rev Microbiol* 2012;10:203-212.
- Delsuc, F., Brinkmann, H. & Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* **6**, 361–375 (2005).
- Eichinger, L. *et al.* The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**, 43–57 (2005).
- Foerster, K. U., von Mering, C., Hooper, S. D. & Bork, P. Environments shape the nucleotide composition of genomes. *EMBO Rep* **6**, 1208–1213 (2005).
- Gadagkar, S. R., Rosenberg, M. S. & Kumar, S. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J. Exp. Zool. B Mol. Dev. Evol.* **304**, 64–74 (2005).
- Gill, E. E. & Brinkman, F. S. L. The proportional lack of archaeal pathogens: Do viruses/phages hold the key? *BioEssays* **33**, 248–254 (2011)..
- Gruber, T. M. & Gross, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Review of Microbiology* **57**, 441–466 (2003).
- Hildebrand, F., Meyer, A. & Eyre-Walker, A. Evidence of Selection upon Genomic GC-Content in Bacteria. *PLoS Genet* **6**, e1001107 (2010).
- Hershberg, R. & Petrov, D. A. Evidence That Mutation Is Universally Biased towards AT in Bacteria. *PLoS Genet* **6**, e1001115 (2010).
- Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Research* **21**, 1552–1560 (2011).
- Ishihama, A. Functional Modulation of Escherichia Coli Rna Polymerase. *Annual Review of Microbiology* **54**, 499–518 (2000).

- Johansson, L., Gafvelin, G. & Arnér, E. S. J. Selenocysteine in proteins—properties and biotechnological use. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1726**, 1–13 (2005).
- Klappenbach JA, Dunbar JM, Schmidt TM. rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* 2000;66:1328-1333.
- Kolaczowski, B. & Thornton, J. W. Long-Branch Attraction Bias and Inconsistency in Bayesian Phylogenetics. *PLoS One* 4, (2009).
- Kupczok, A., Schmidt, H. A. & Von Haeseler, A. Accuracy of phylogeny reconstruction methods combining overlapping gene data sets. *Algorithms Mol Biol* **5**, 37 (2010).
- Lafay, B., Atherton, J. C. & Sharp, P. M. Absence of translationally selected synonymous codon usage bias in *Helicobacter pylori*. *Microbiology* **146**, 851–860 (2000).
- Lightfield, J., Fram, N. R. & Ely, B. Across Bacterial Phyla, Distantly-Related Genomes with Similar Genomic GC Content Have Similar Patterns of Amino Acid Usage. *PLoS ONE* **6**, e17677 (2011).
- McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Origin of an Alternative Genetic Code in the Extremely Small and GC-Rich Genome of a Bacterial Symbiont. *PLoS Genet* **5**, e1000565 (2009).
- McDaniel, L. D. *et al.* High Frequency of Horizontal Gene Transfer in the Oceans. *Science* **330**, 50–50 (2010).
- Lang, J. M., Darling, A. E. & Eisen, J. A. Phylogeny of Bacterial and Archaeal Genomes Using Conserved Genes: Supertrees and Supermatrices. *PLoS ONE* **8**, e62510 (2013).
- Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
- Ponting, C. P., Aravind, L., Schultz, J., Bork, P. & Koonin, E. V. Eukaryotic Signalling Domain Homologues in Archaea and Bacteria. Ancient Ancestry and Horizontal Gene Transfer. *Journal of Molecular Biology* **289**, 729–745 (1999).
- Philippe, H. *et al.* Phylogenomics of Eukaryotes: Impact of Missing Data on Large Alignments. *Mol Biol Evol* **21**, 1740–1752 (2004).
- Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**, e9490 (2010).
- Rastogi, G. *et al.* Microbial and Mineralogical Characterizations of Soils Collected from the Deep Biosphere of the Former Homestake Gold Mine, South Dakota. *Microbial Ecology* **60**, 539–550 (2010).
- Recolons, C. A. *et al.* High-Resolution Crystal Structure of an Outer Membrane-Anchored Endolytic Peptidoglycan Lytic Transglycosylase (MltE) from *Escherichia coli*. *Biochemistry* **50**, 2384–2386 (2011).
- Scheurwater, E., Reid, C. W. & Clarke, A. J. Lytic transglycosylases: Bacterial space-making autolysins. *The International Journal of Biochemistry & Cell Biology* **40**, 586–591 (2008).
- Shafaat, H. S., Weber, K., Petrenko, T., Neese, F. & Lubitz, W. Key hydride vibrational modes in [NiFe] hydrogenase model compounds studied by resonance Raman spectroscopy and density functional calculations. *Inorg Chem* **51**, 11787–11797 (2012).
- Shand, R. F. & Leyva, K. J. in *Bacteriocins* (Riley, M. A. & Chavan, M. A.) 93–109 (Springer Berlin Heidelberg, 2007). at <http://www.springerlink.com/content/nr538158x2612627/abstract/>
- Stamatakis, A. *et al.* RAxML-Light: A Tool for computing TeraByte Phylogenies. *Bioinformatics* (2012). doi:10.1093/bioinformatics/bts309

Sun D, Lee G, Lee JH, Kim HY, Rhee HW, Park SY, Kim KJ, Kim Y, Kim BY, Hong JI, Park C, Choy HE, Kim JH, Jeon YH, Chung J. A metazoan ortholog of SpoT hydrolyzes ppGpp and functions in starvation responses. *Nat Struct Mol Biol* 2010;17:1188-1194.

Thomm, M.: Transcription: Mechanism and Regulation. In: *Archaea Molecular and Cellular Biology*. R. Cavicchioli (ed.). ASM Press, Washington DC, U.S.A., Chapter 6, 139-157 (2007).

Tozawa Y, Nomura Y. Signalling by the global regulatory molecule ppGpp in bacteria and chloroplasts of land plants. *Plant Biol (Stuttg)* 2011;13:699-709

Valentine DL. Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat Rev Microbiol* 2007;5:316-323.

Woolfit, M., Iturbe-Ormaetxe, I., McGraw, E. A. & O'Neill, S. L. An Ancient Horizontal Gene Transfer between Mosquito and the Endosymbiotic Bacterium *Wolbachia pipientis*. *Mol Biol Evol* **26**, 367–374 (2009).

Wu, M. & Eisen, J. A. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* **9**, R151 (2008).

Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28**, 1033–1034 (2012).

Wu, J. & Xie, J. Magic spot: (p) ppGpp. *Journal of Cellular Physiology* **220**, 297–302 (2009).

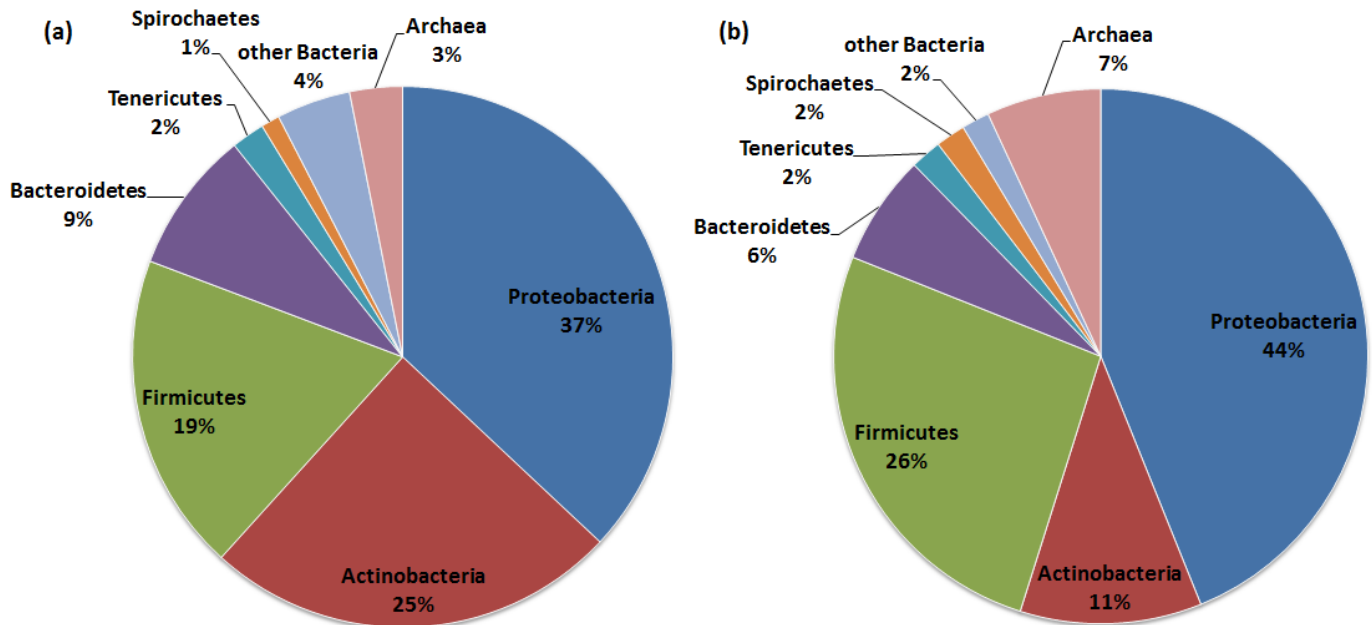
Wu, H., Zhang, Z., Hu, S. & Yu, J. On the molecular mechanism of GC content variation among eubacterial genomes. *Biol Direct* **7**, 2 (2012).

Wu, D. *et al.* A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**, 1056–1060 (2009).

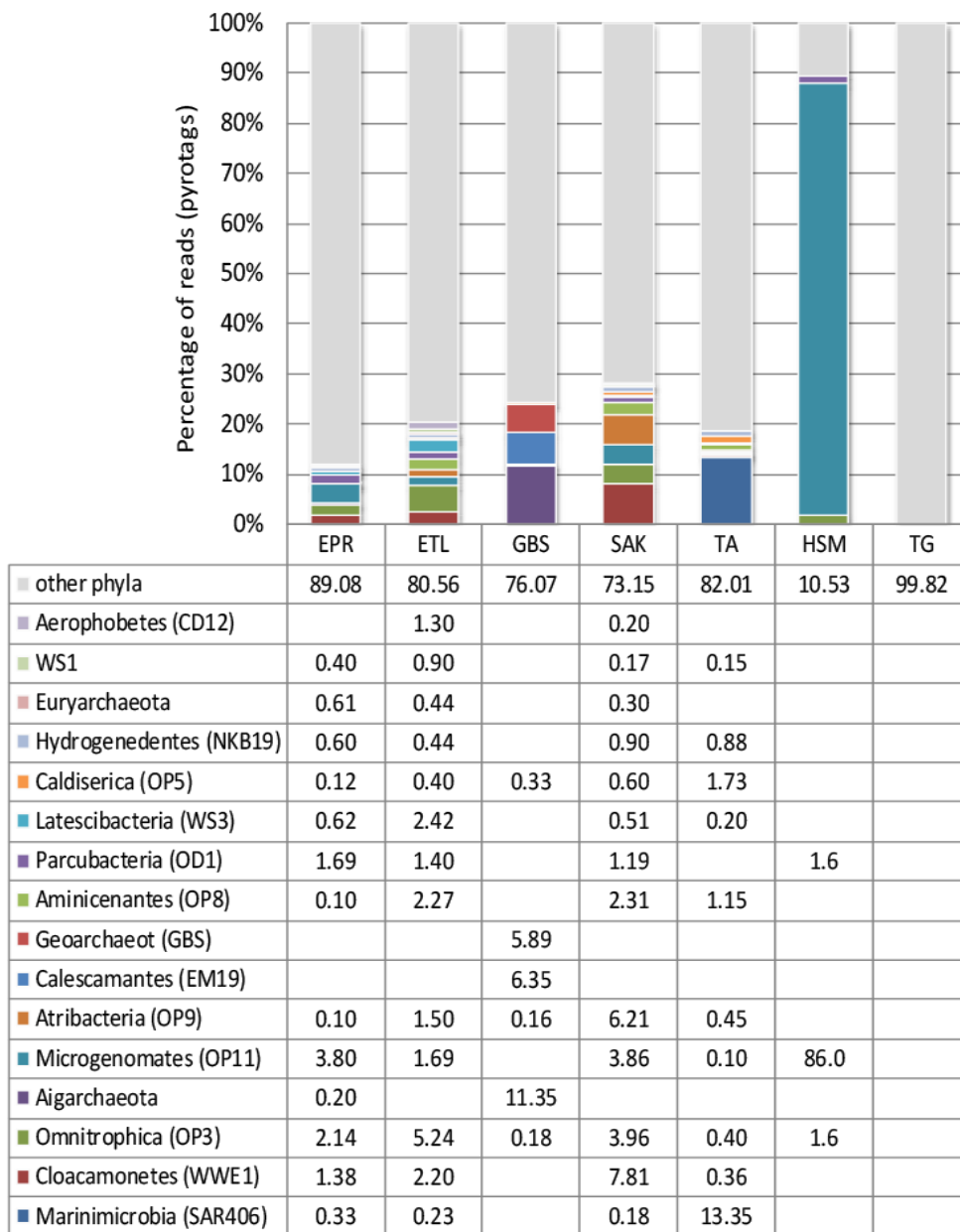
Yang, Z. & Rannala, B. Molecular phylogenetics: principles and practice. *Nat Rev Genet* **13**, 303–314 (2012).

Yamano, F. *et al.* UGA is read as tryptophan in *Mycoplasma capricolum*. *Proc Natl Acad Sci U S A* **82**, 2306–2309 (1985).

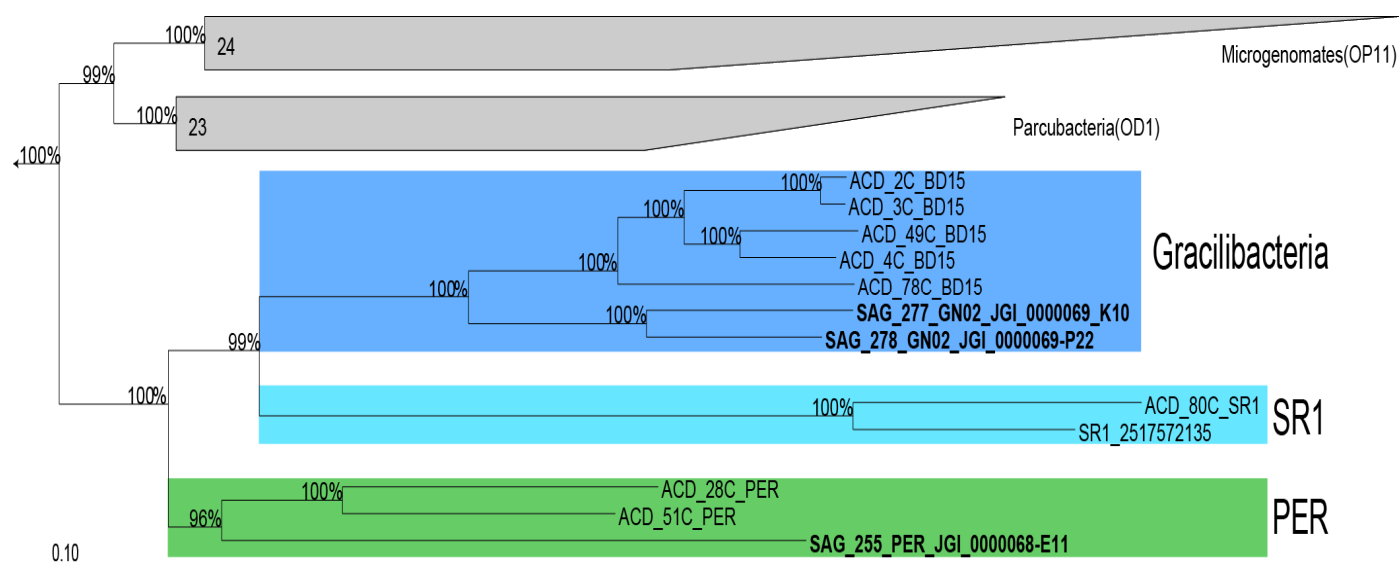
Supplementary Figures



Supplementary Figure 1 | Culturing and sequencing bias towards four bacterial phyla. (a) Phylum-level distribution of bacterial isolates, including isolates from the domain Archaea. The data are based on the DSMZ-German Collection of Microorganisms and Cell Cultures (www.dsmz.de) database and show an overrepresentation of the four bacterial phyla Proteobacteria, Actinobacteria, Firmicutes, and Bacteroidetes in culture collections. The four phyla comprise more than 88% of all isolates maintained at DSMZ. Note that DSMZ does not cultivate Cyanobacteria. These findings are consistent with a survey based on the Australian Collection of Microorganisms (Hugenholtz 2002). **(b)** Sequenced bacterial and archaeal genomes. Data were obtained from the Integrated Microbial Genomes (IMG) system (img.jgi.doe.gov) on December 2012. The total percentage of the four most sequenced phyla is 87.7%.



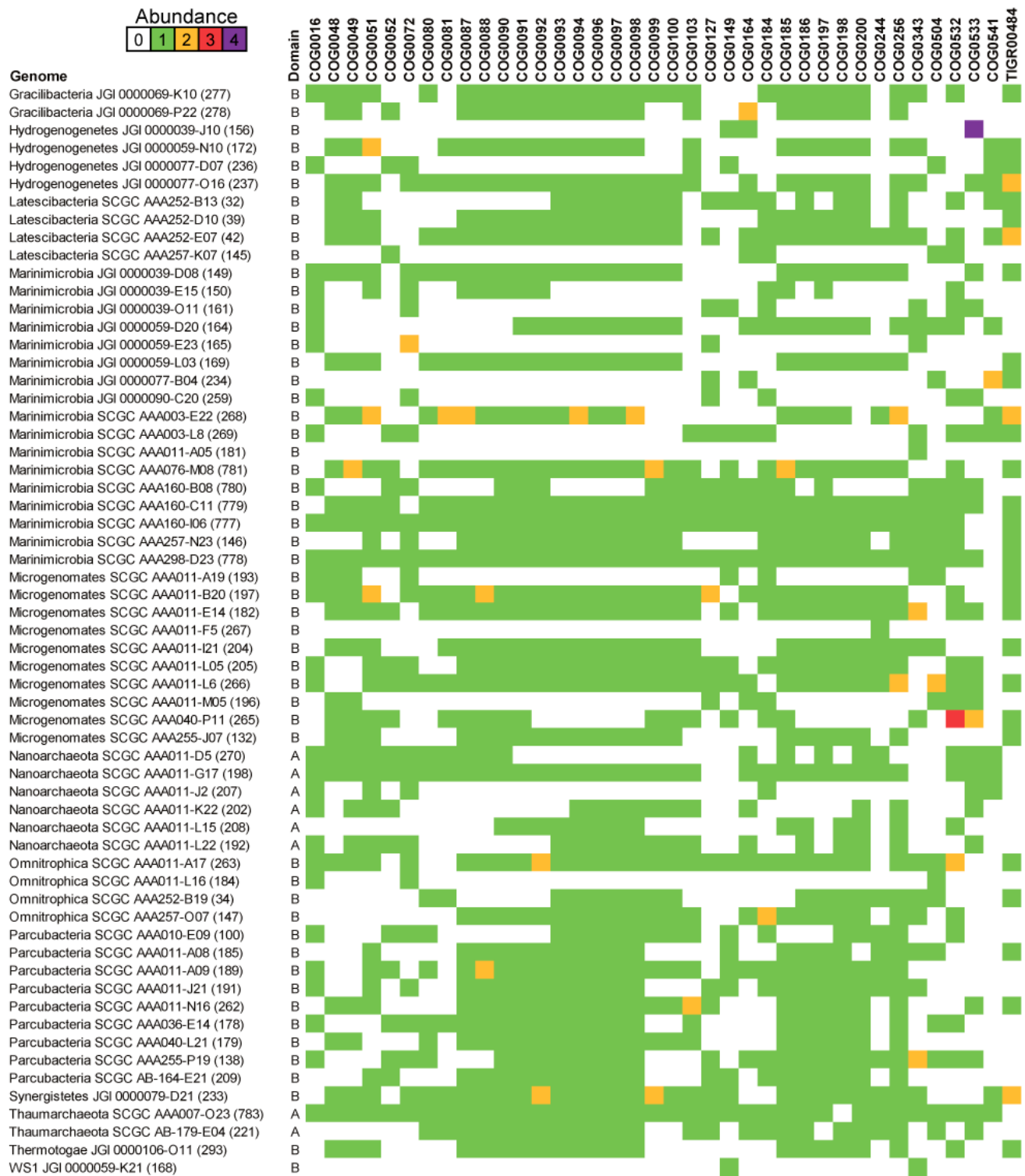
Supplementary Figure 2 | Microbial community profile based on SSU rRNA gene sequences. The phylum-level classifications, including all phyla with equal to or greater than 1% of the total number of reads (pyrotags) analysed by PyroTagger (Kunin and Hugenholtz, 2010). Percentage of reads in the bar chart represent percentages given in the table below the chart. Candidate phyla investigated in this study are highlighted in color, whereas known phyla and other candidate phyla are displayed in light gray. The domain is provided after the phylum name: (A) for Archaea and (B) for Bacteria. The sampling site abbreviations are EPR (East Pacific Rise), ETL (Etoliko Lagoon), GBS (Great Boiling Spring), SAK (Sakinaw Lake), TA (Terephthalate degrading reactor), HSM (Homestake Mine), TG (Tropical Gyre in the South Atlantic). The HSM sample contained a very low number of cells ($\sim 10^3$ cells/ml) and had to be amplified by Multiple Displacement Amplification (MDA), which most likely introduced a bias towards certain phyla. In the TG sample, candidate phyla were minor representatives of the microbial community as based on the community profile, yet we recovered two Marinimicrobia (SAR406) single amplified genomes (SAGs) from this environment. For two of the sampling sites, Gulf of Maine (GOM) and Hawaii Ocean Time-series (HOT), no SSU rRNA gene data were generated due to insufficient DNA.



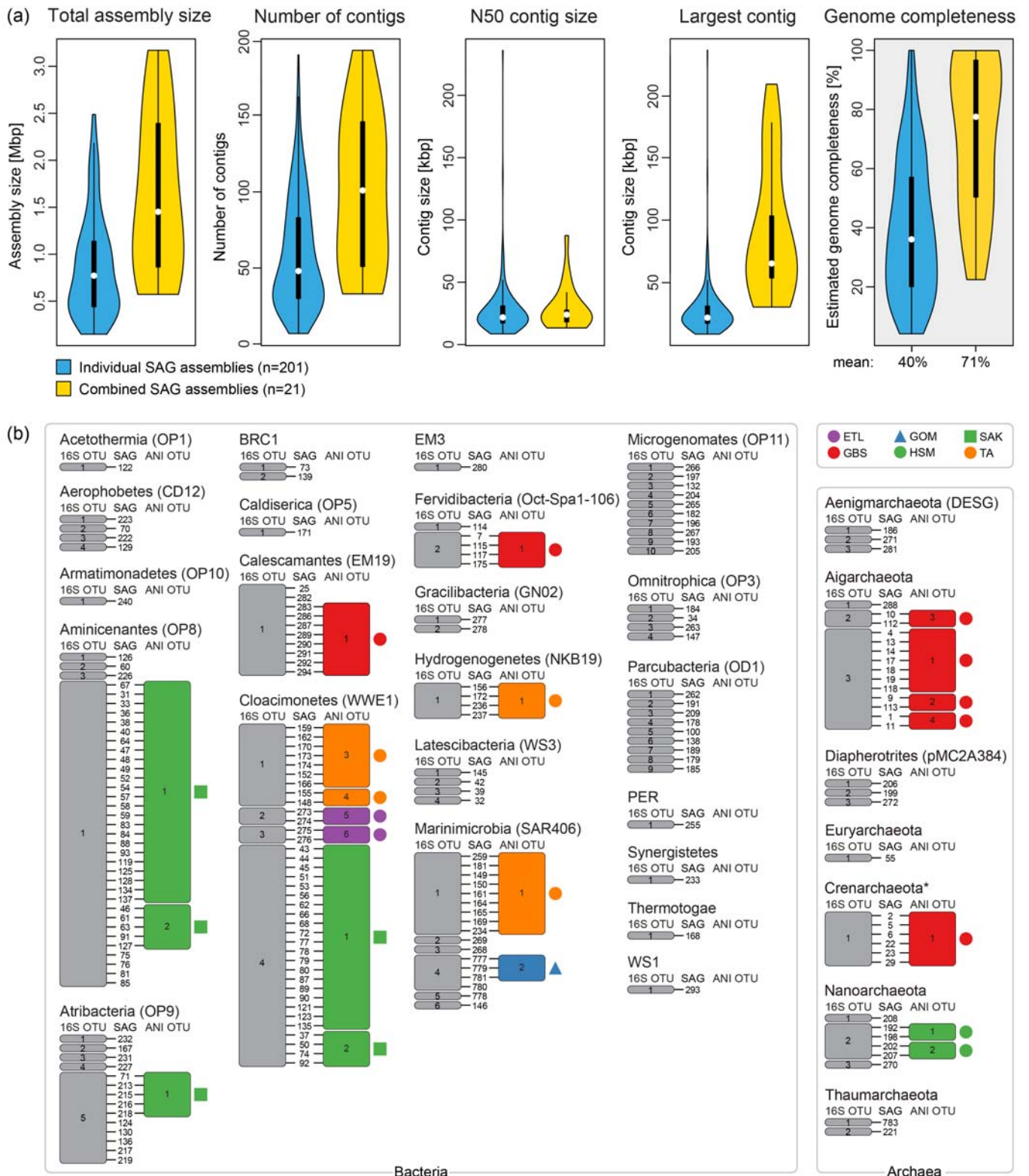
Supplementary Figure 3 | Phylogenetic inference of the Gracilibacteria and the PER group. Phylogenetic maximum likelihood tree (fasttree, JTT, CAT) of Gracilibacteria, the PER group, the very recently published SR1 genomes, and related phyla based on up to 38 marker genes. The scale bar represents 10% estimated sequence divergence.





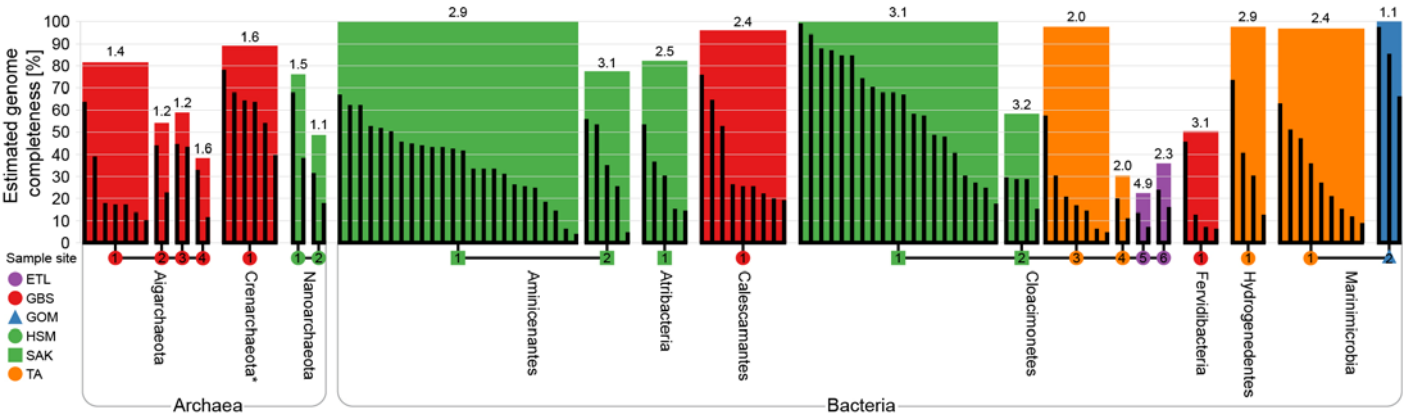


Supplementary Figure 4 | Fidelity matrix of phylogenetic marker genes. Absence, presence and number of 38 phylogenetic marker genes (x-axes) found in the 201 SAGs (y-axes). Green indicates that one copy of the marker gene is present, while yellow, red and purple show that two, three and four copies, respectively, are present. White indicates that the gene was not annotated. For a detailed list of the 38 marker genes see Supplementary Table 2. Please note that COG0482 comprises translation elongation factors (GTPases) for which gene duplication events are known.

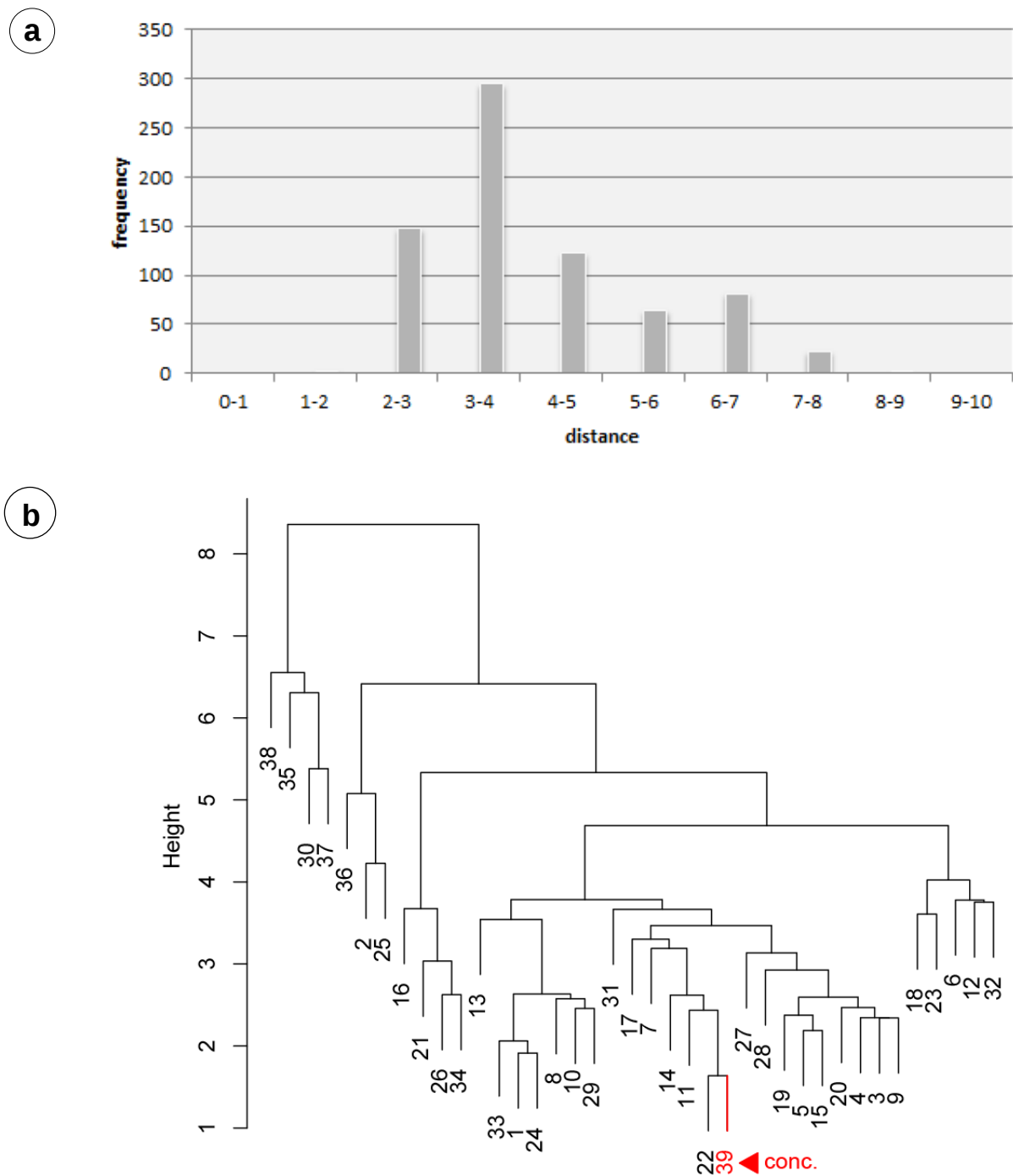


Supplementary Figure 5 | Assembly statistics and combined assembly lineup. (a) Assembly statistics are shown for of single cells (blue) and for combined assemblies (yellow). The violin plots show from left to right the total assembly size, number of contigs, N50 contig size, and the largest contig. The plots with gray background on the right-hand side show the estimated genome completeness based on

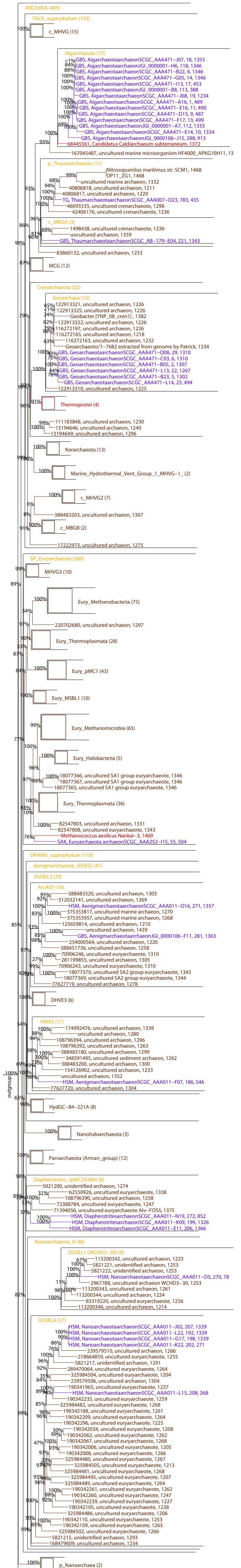
presence/absence of 139 bacterial and 162 archaeal conserved, single copy genes in 90% of all genomes, marker genes. Black boxes spans the region between the 25th and 75th quartile, the white dot indicates the median. **(b)** SSU rRNA gene OTUs vs ANI. SAGs for which SSU rRNA gene sequences clustered within 97% OTUs were used for the analysis of the Average Nucleotide Identity (ANI). In cases where the ANI among SAGs was equal or greater than 97% the genome sequences were used for combined assemblies (colored boxes).



Supplementary Figure 6 | Estimated genome completeness of co-assembled single amplified genomes. Shown is the estimated genome completeness of the co-assemblies (colored rectangles) and of the individual SAGs used for the co-assemblies (black bars). Combined assemblies were generated for SAGs derived from the same environment that yielded an average nucleotide identity (ANI) of at least 97% and were within the same operational taxonomic unit (OTU) defined by SSU rRNA gene identity. The color of each combined assembly (colored rectangle) and the symbols below the bars correspond to the sampling sites shown in Fig. 1. Numbers in the symbols indicate different ANI-groups of the same phylum. Numbers above the rectangles indicate the estimated complete size of the co-assembled genomes in Mbp. The assemblies are ordered alphabetically by domain, phylum, and ANI-group. Detailed genome statistics for each single cell are provided in Supplementary Table * SAGs belong to the subgroup Geoarchaeota within the Crenarchaeota.

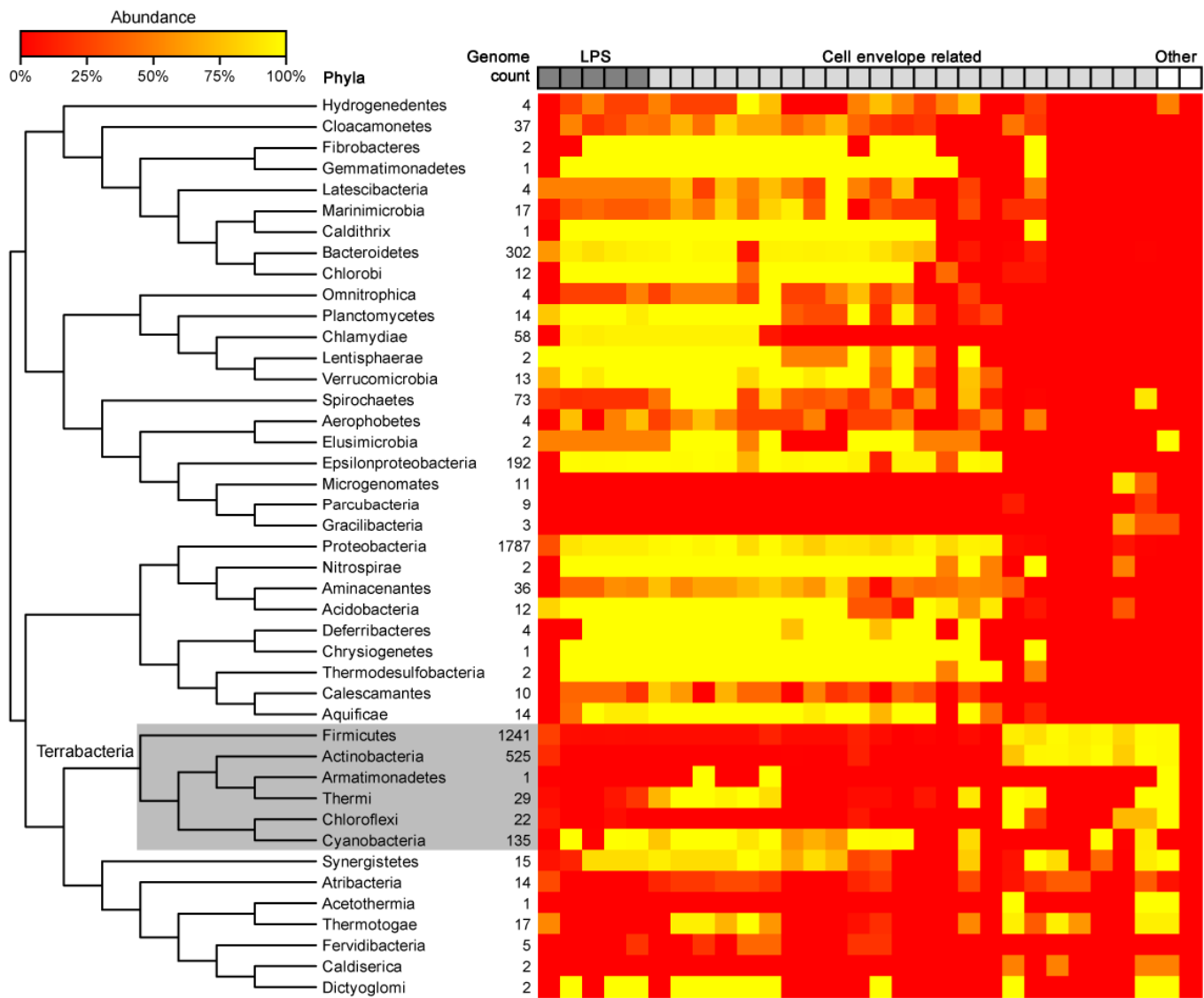


Supplementary Figure 7 | Congruence analysis of individual gene trees. The pairwise geodesic distances were calculated between 39 maximum likelihood (FastTree2) marker gene trees. These included 38 individual gene trees as well as the concatenated alignment tree of all 38 genes. **(a)** The frequency of occurrences is plotted against the pairwise geodesic distances. No outliers were detected, which suggests that all individual gene trees represent comparable branch lengths and topologies. **(b)** The hierarchical clustering of pairwise geodesic distances shows a congruent clustering pattern with four trees (29, 34, 36, and 37) clustering slightly removed from the rest. The concatenated 38 marker gene tree (conc.; highlighted with a red arrow) clusters with the congruent individual gene trees, which is consistent with an additive phylogenetic signal of the individual gene trees.

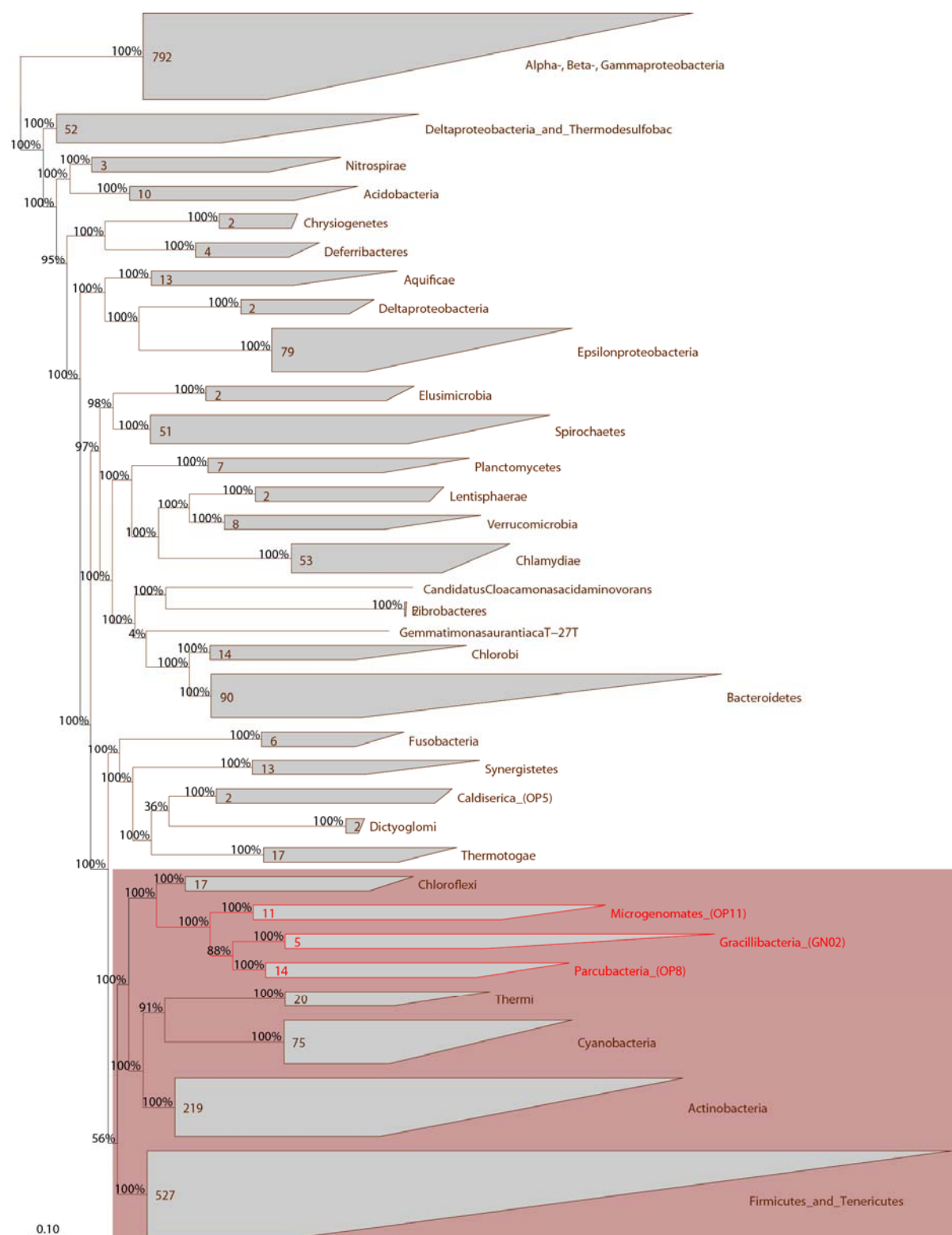




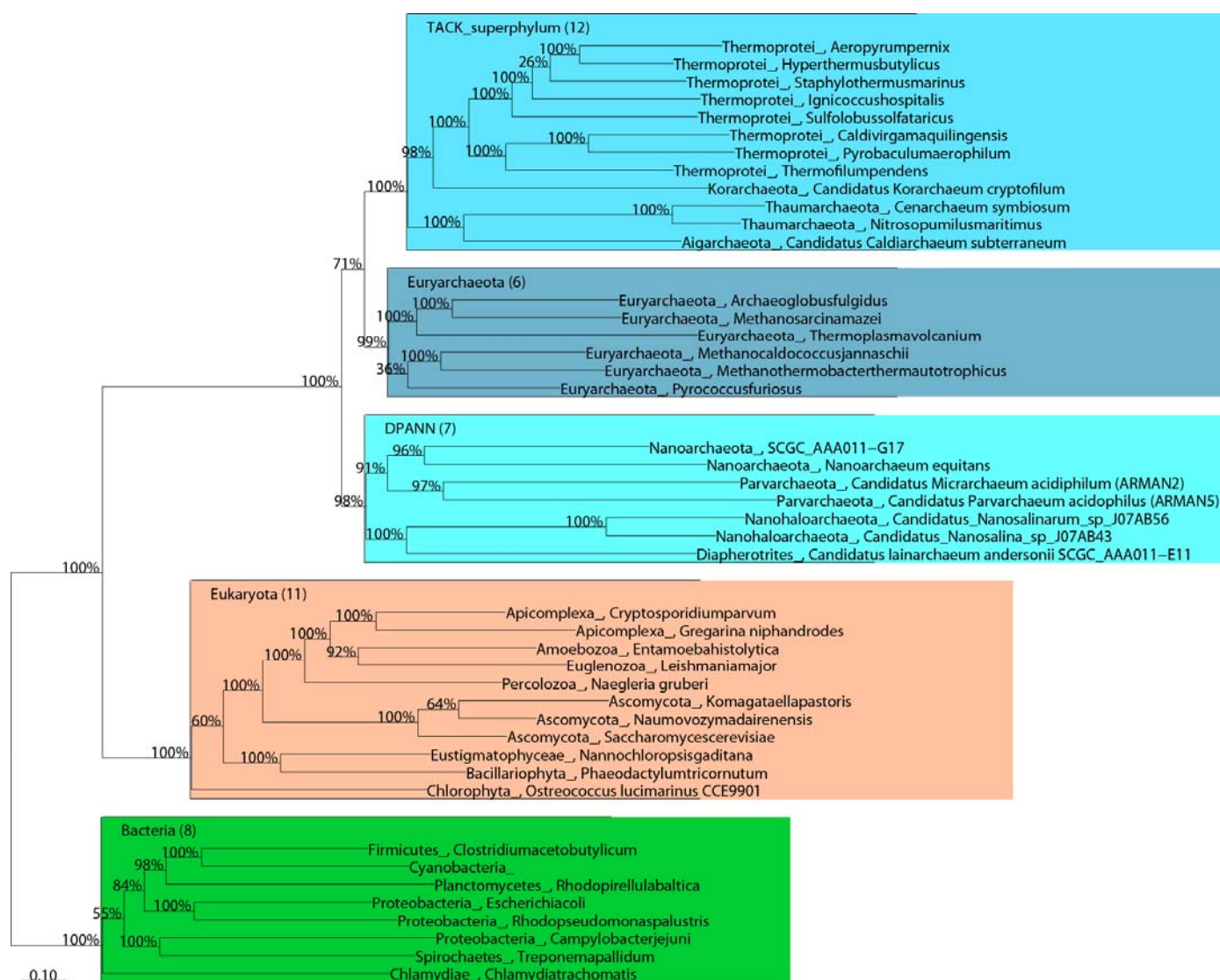
Supplementary Figure 8 | SSU rRNA gene maximum likelihood phylogenetic inference. (a) domain Archaea; **(b)** domain Bacteria. Sequences are grouped on the phylum level, except for the phylum Proteobacteria, which is polyphyletic in our analysis. Groups containing SSU rRNA genes from single-cell genomes assembled in this study are highlighted in orange, and the species are highlighted in blue.



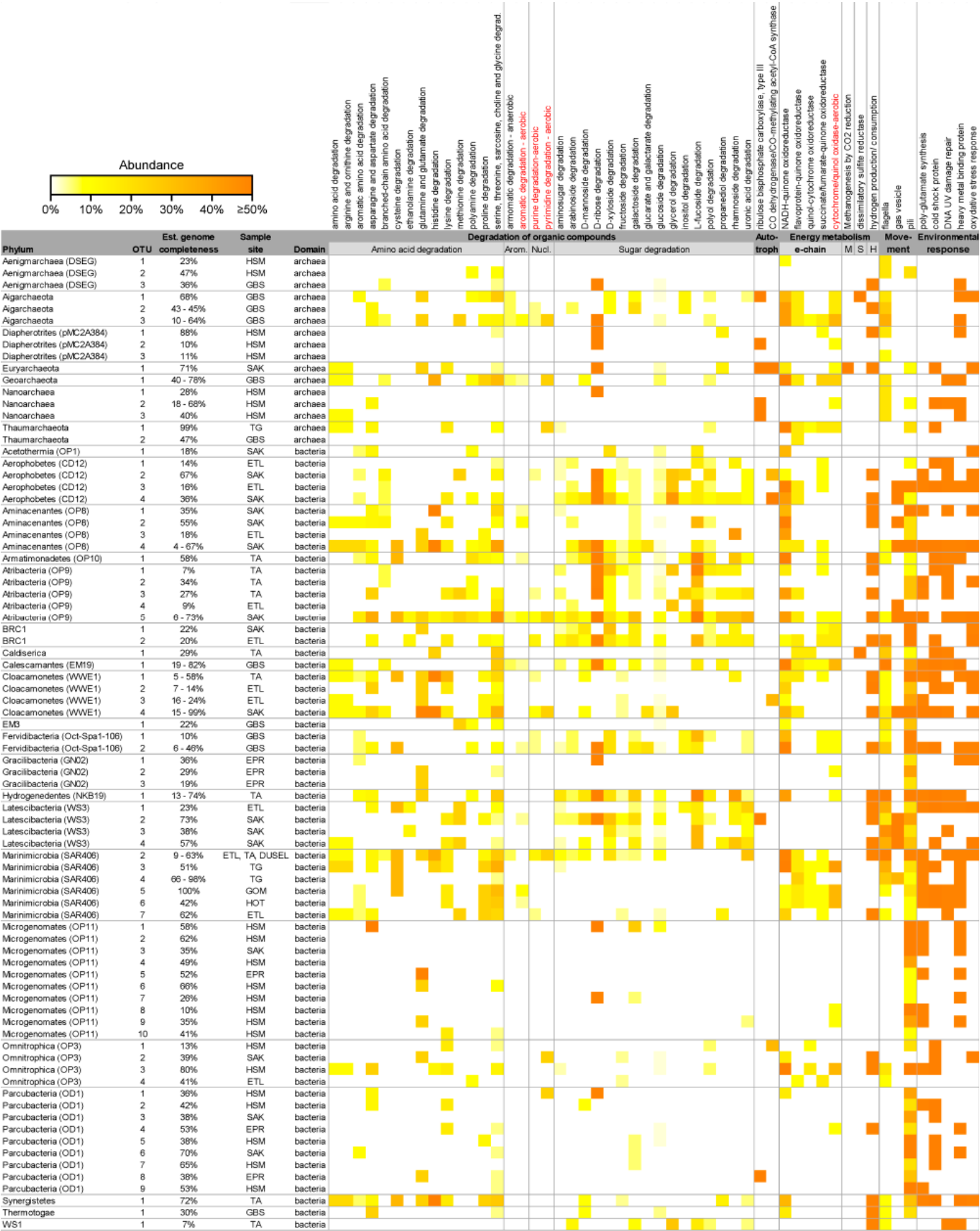
Supplementary Figure 9 | Bacterial cell envelope architecture. Comparison of the average abundance of pfams involved in cell envelope architecture per phylum. The percentage of genomes annotated with a particular pfam is shown as a heatmap from red (low) to high (yellow). The superphylum Terrabacteria is highlighted by a gray box. The superphylum Patescibacteria is represented by all three phyla Microgenomates, Parcubacteria, and Gracilibacteria. LPS=Lipopolysaccharides.



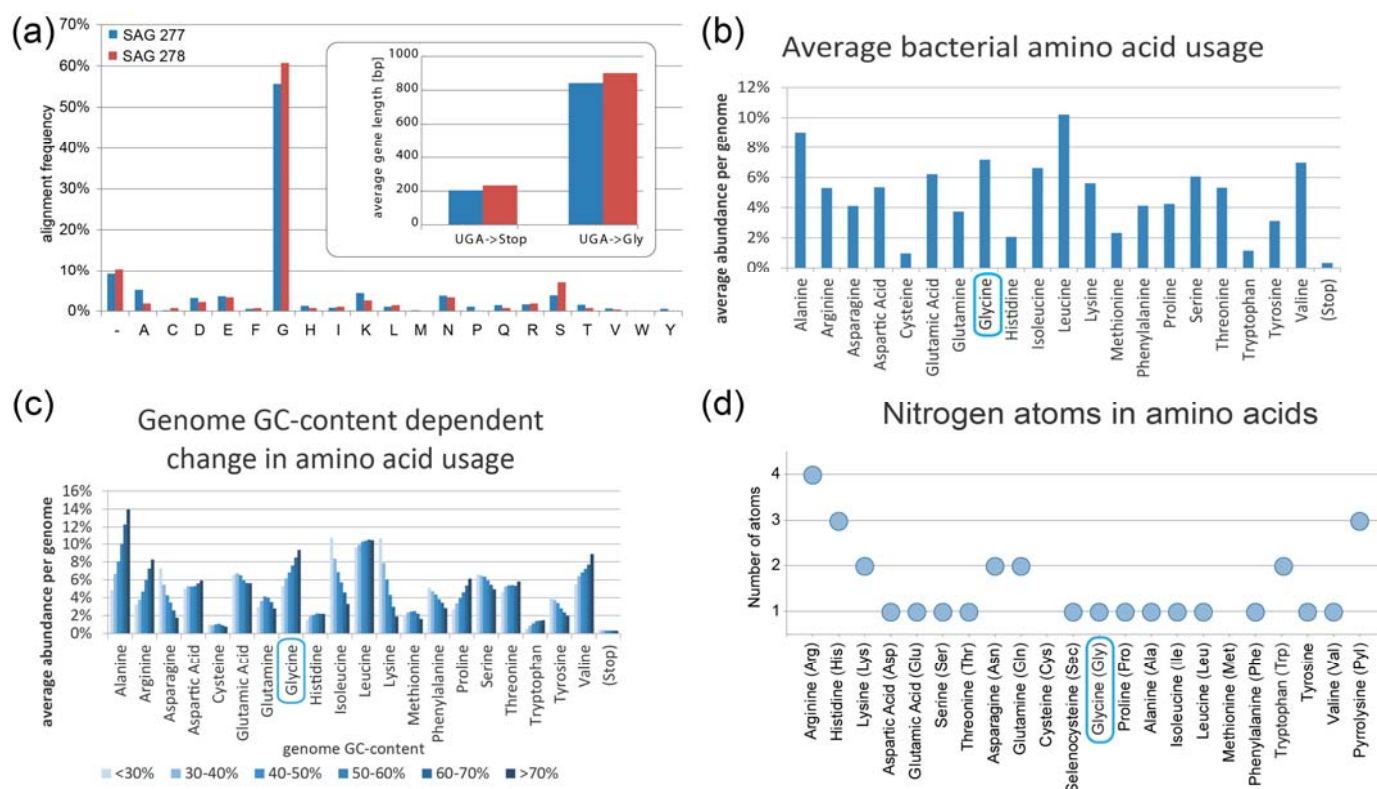
Supplementary Figure 10 | Phylogeny of Patescibacteria and Terrabacteria. Phylogenetic maximum likelihood tree (fasttree, JTT, CAT) based on a concatenated alignment of an extended set of up to 83 marker genes with focus on the Patescibacteria. The superphylum Terrabacteria is highlighted by a red box and the three members of the Patescibacteria are highlighted with a red font.



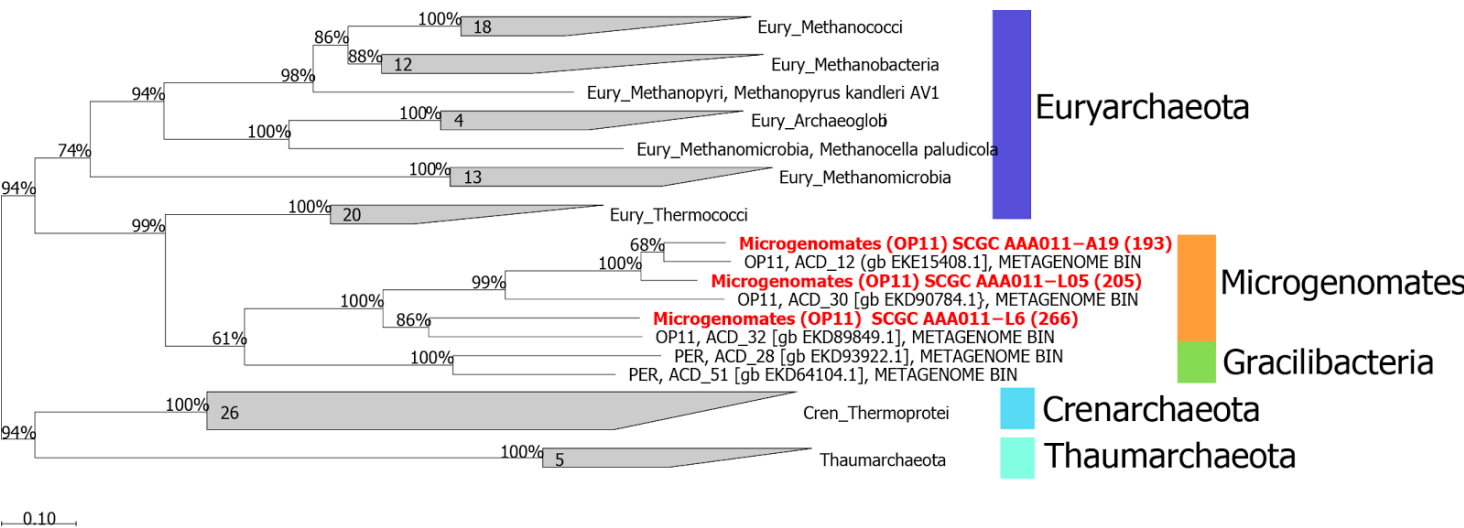
Supplementary Figure 11 | Phylogeny of Archaea, Bacteria and Eukaryotes. Phylogenetic maximum likelihood tree (fasttree, JTT, CAT) based on a concatenated alignment of up to 38 marker genes. Shown are taxa from Bacteria, Eukaryota, and Archaea. The latter include taxa from the proposed superphylum DPANN. The topology and 100% bootstrap values (bootstrap is the standard method to estimate reliability) support the three domains of life.



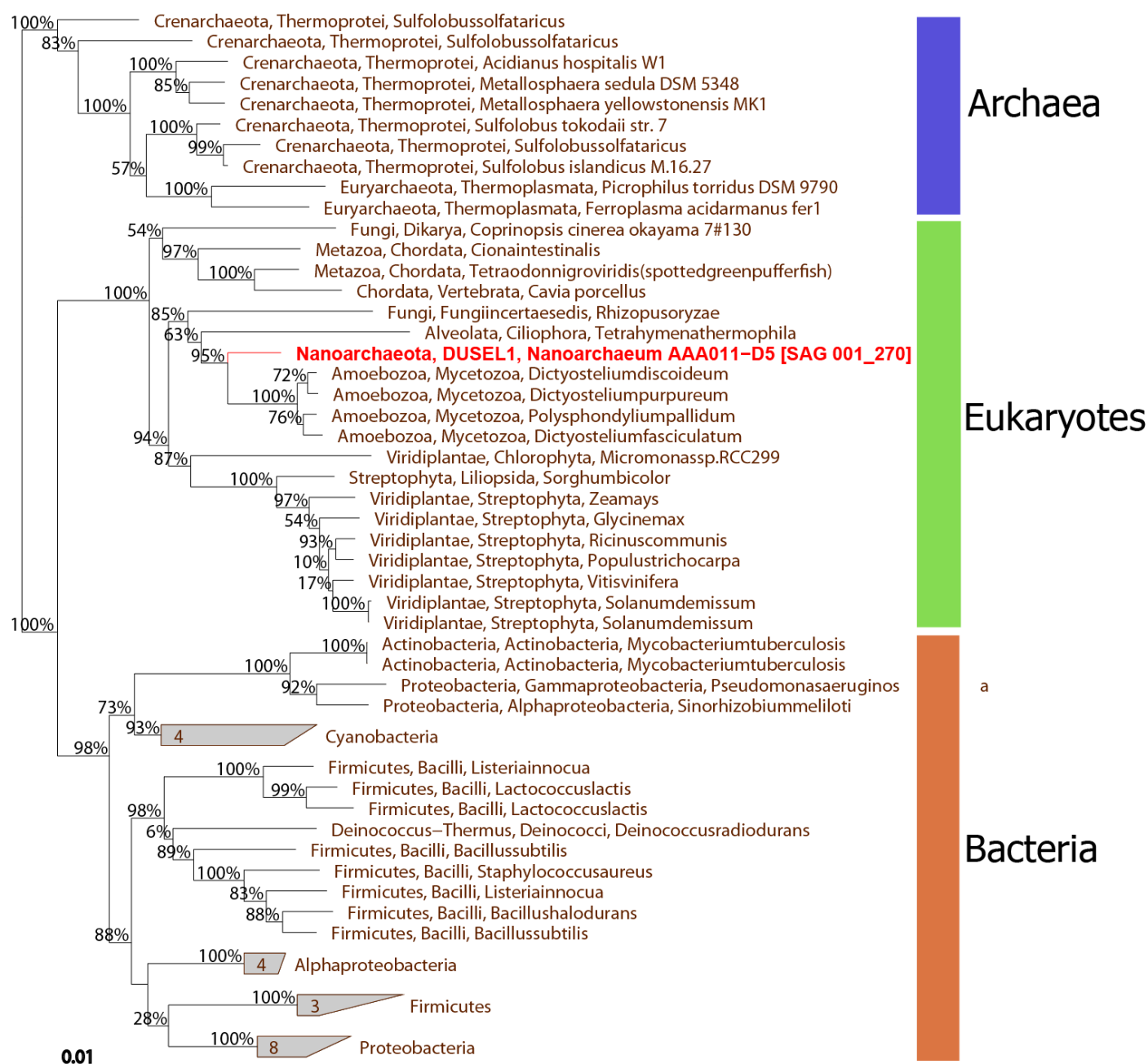
Supplementary Figure 12 | Functional diversity inferred from the MDM single-cell genomes. Single-cell genomes are grouped by phylum and OTUs. The major functions encoded in the annotated single-cell genomes are grouped into functional categories in the header and are shown on the x-axis against the single-cell genome SSU rRNA OTUs on the y-axis. The presence of genes encoding proteins of each functional category is given as a heat map ranging from white (no gene present) to bright yellow (10% of the genes present) to dark orange (more than 50% of the genes present).



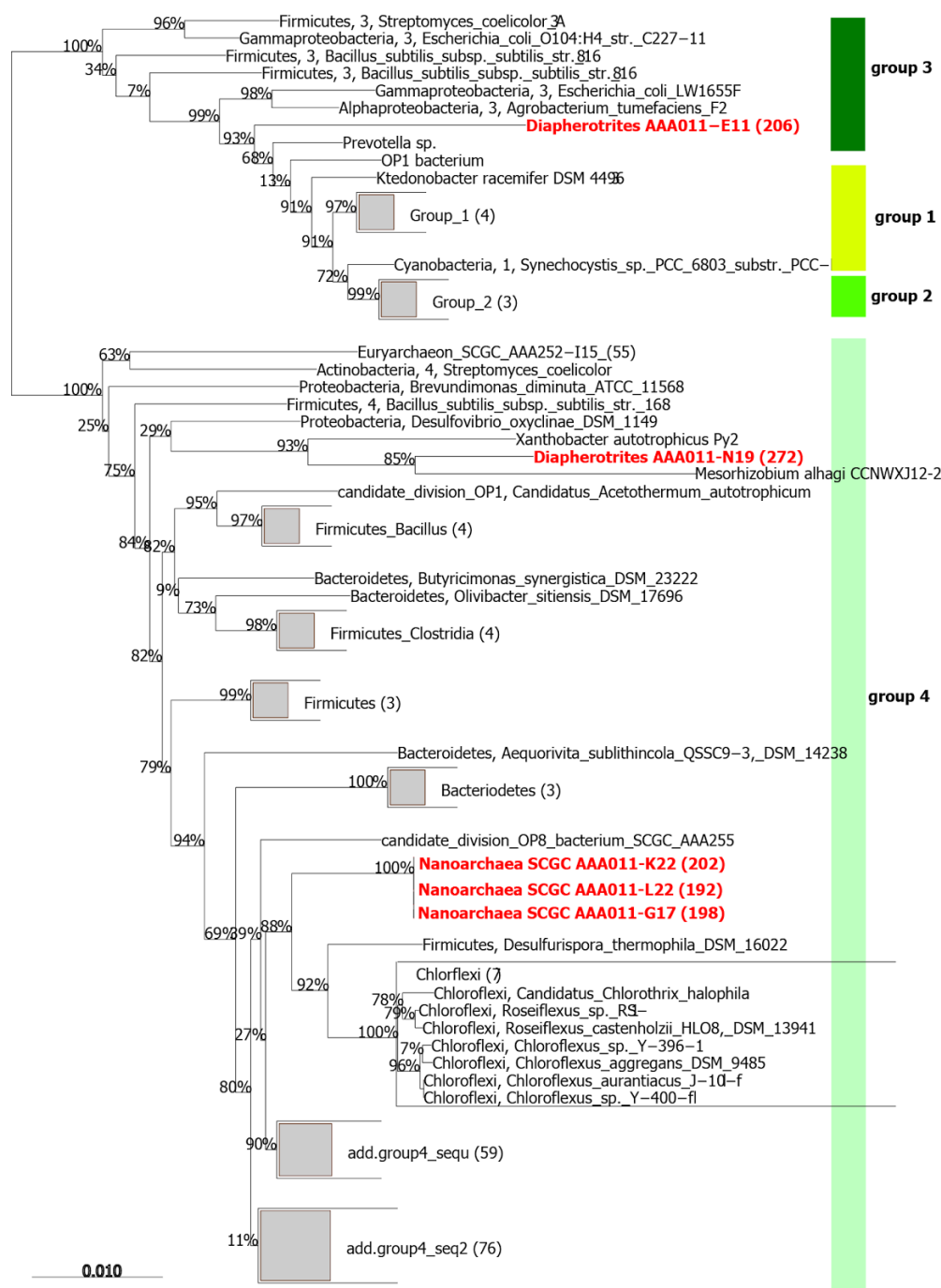
Supplementary Figure 13 | Predicted codon and amino acid usage of Gracilibacteria (GNO2) genomes. (a) UGA stop-codon encodes glycine in two single-cell genomes. Shown are amino acid transition frequencies from UGA-stop-codons for all predicted genes of SAG 277 and SAG 278 derived from protein sequence alignments with best hits to the NCBI non-redundant protein database. Gaps are shown as “-“. Nested figure: Differences in the average gene lengths after gene prediction using standard translation table 11 (UGA->Stop) and a custom translation table (UGA->Gly). (b) Glycine is the 3rd most used amino acid among bacterial genomes. All available finished genomes in IMG (version 4.0.1 December 2012) from Bacteria (1962 genomes total) were included in this and the following charts of this figure. (c) Amino acid use correlates with genomic GC content. Shown are all finished bacterial genomes currently available in IMG (see above). (d) Nitrogen atoms in the 22 amino acids. Glycine is highlighted with a blue frame.



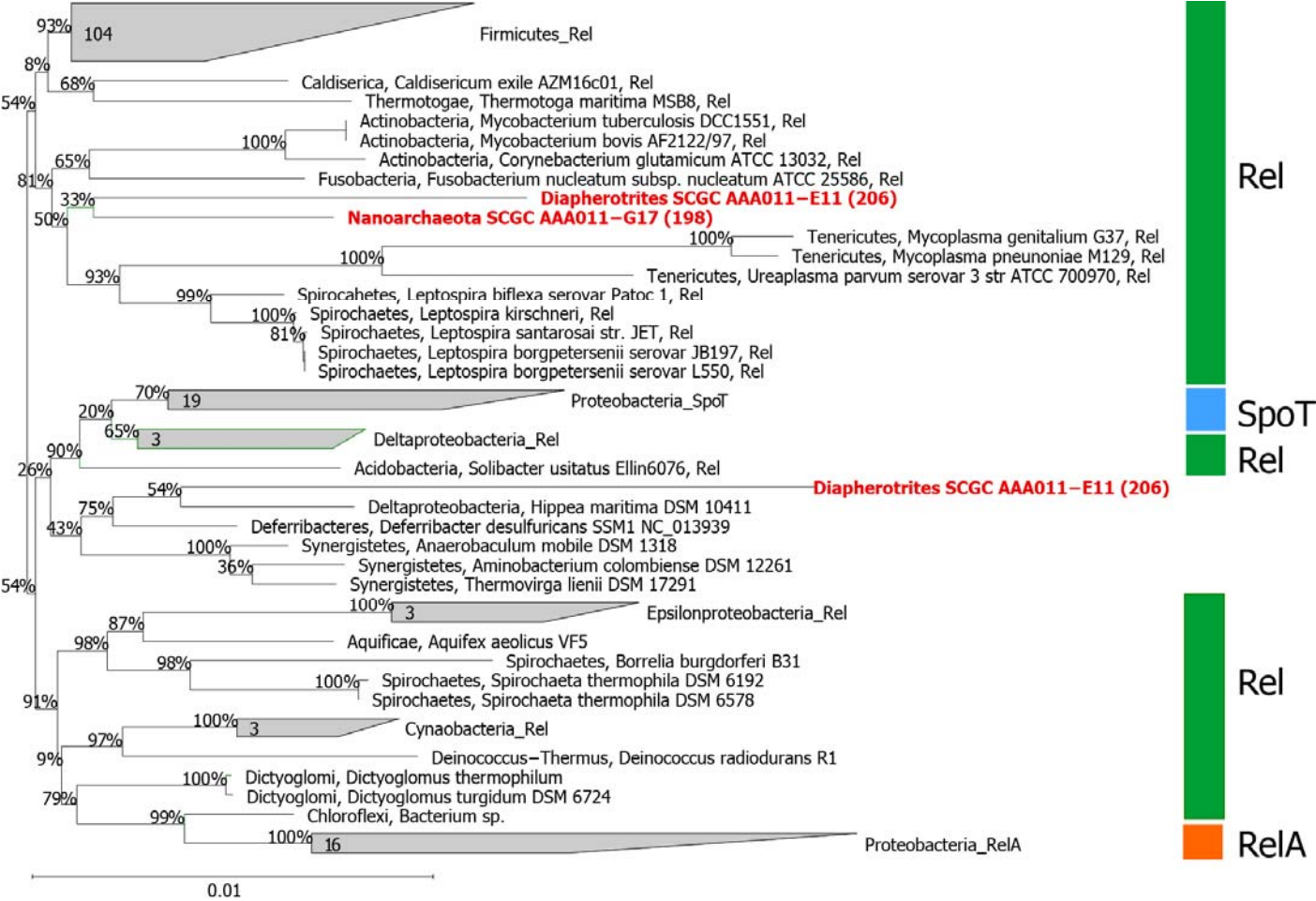
Supplementary Figure 14 | Archaeal PurP genes in Patescibacteria. Phylogenetic maximum likelihood tree (fasttree, JTT, CAT) of PurP genes in Microgenomates (OP11) and Gracilibacteria (GN02). Genes in SAG genomes from this study are highlighted in red. The scale bar represents 10% estimated sequence divergence.



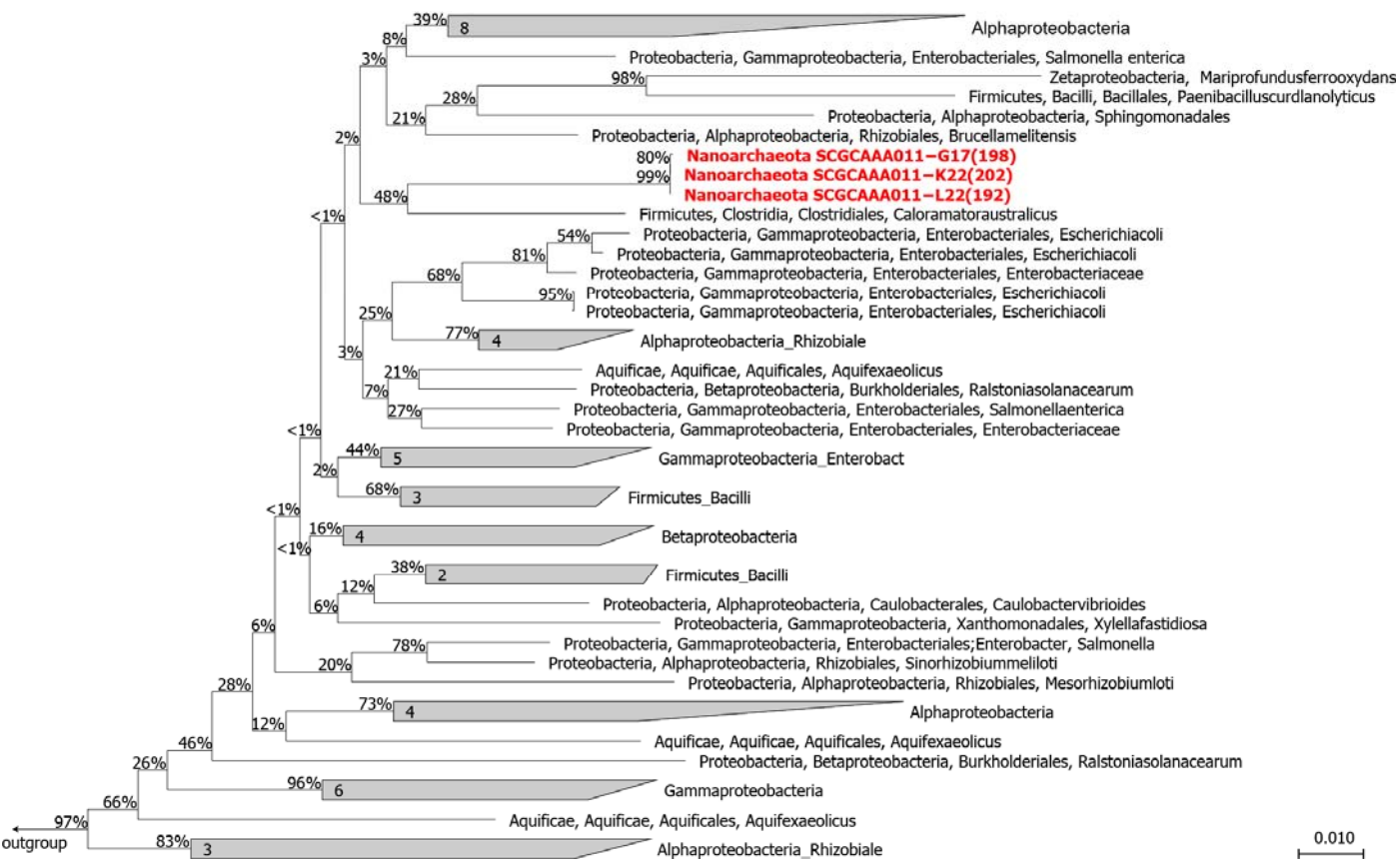
Supplementary Figure 15 | Lateral gene transfer from an Eukaryote to a Nanoarchaeum. Phylogenetic maximum likelihood tree (fasttree, JTT, CAT) of 55 genes representing COG1252 (Ndh; NADH dehydrogenase, FAD-containing subunit) spanning the 3 domains of life Archaea, Bacteria and Eukaryota. The Nanoarchaeum genome from this study is highlighted in red. Note that while for the archaeal clade, the maximum diversity available in the NCBI nr database was included, this data set is limited and cannot be used to test the Eocyte theory.



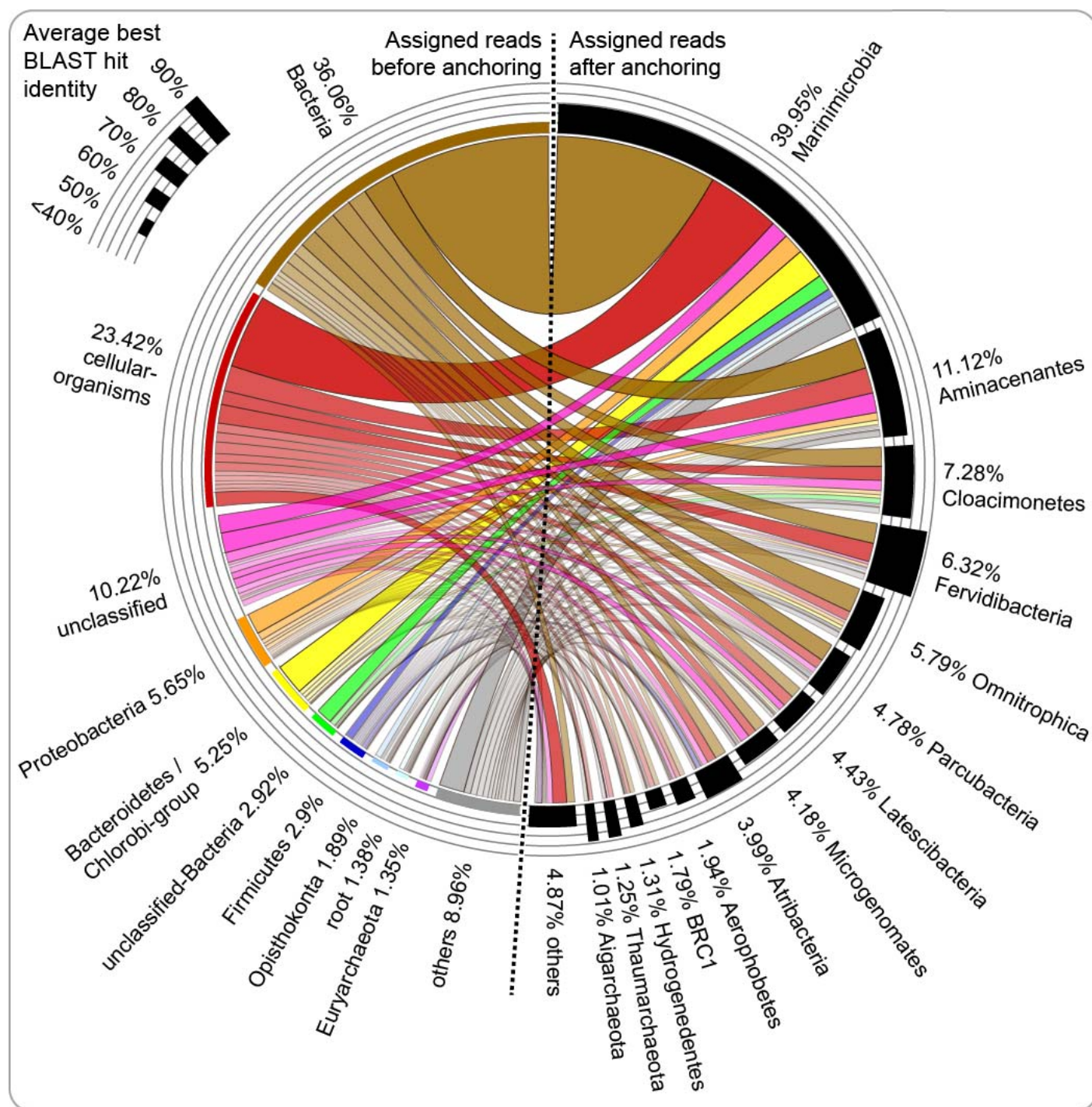
Supplementary Figure 16 | Phylogenetic inference of sigma factor- related proteins. Maximum likelihood tree of conserved domains 2 and 4 of archaeal sigma factor-related proteins within the four main groups of the sigma-70 family. Group 1 contains the sigma-70 (gene *rpoD*), after which the family was named and which functions as primary sigma factor; group 2 contains proteins closely related to group 1 but which are non-essential; group 3 contains specialized sigma factors (e.g. heat shock, sporulation); group 4 (ECF family) includes sigma 24 (the *E. coli* gene *rpoE* sigma factor). Genes in SAG genomes from this study are highlighted in red.



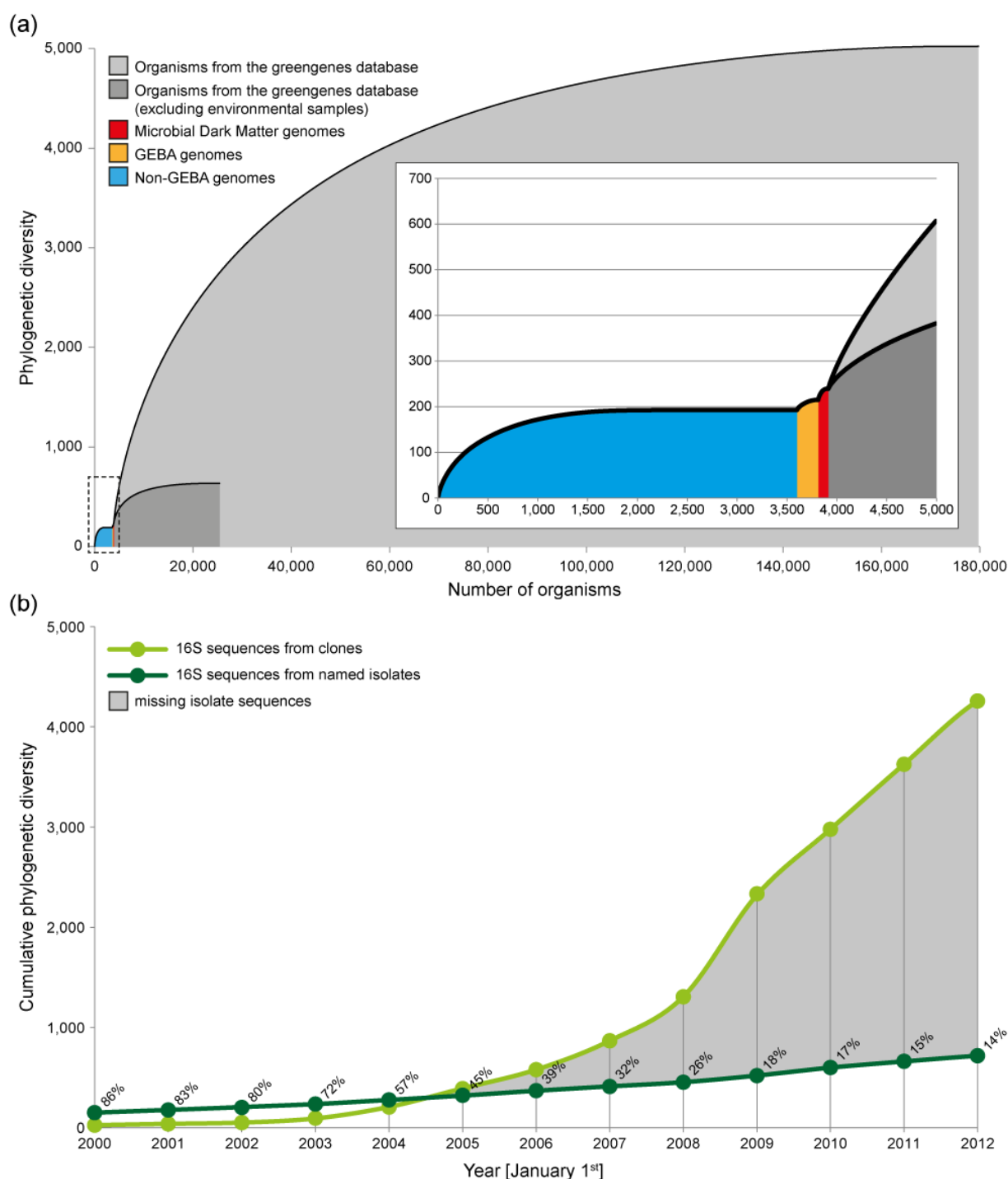
Supplementary Figure 17 | Phylogenetic inference of long RelA/SpoT homologs. Full-length homologs of RelA/SpoT (produce ppGpp in stringent response) having a synthase domain, a hydrolase domain, and a regulatory domain are found in two SAGs (Nanoarchaeota, Diapherotrites). The Diapherotrites genome contains two RelA/SpoT homologs. Subgroup of long RelA/SpoT homologs are RelA and SpoT, named after the RelA and SpoT proteins of *Escherichia coli*, and Rel, named after the ancestral Rel protein found in many bacteria.



Supplementary Figure 18 | Phylogenetic inference of lytic murein transglycosylase genes. Maximum likelihood tree (fasttree, JTT, CAT) showing three SAGs from the Nanoarchaeota encoding lytic murein transglycosylase.



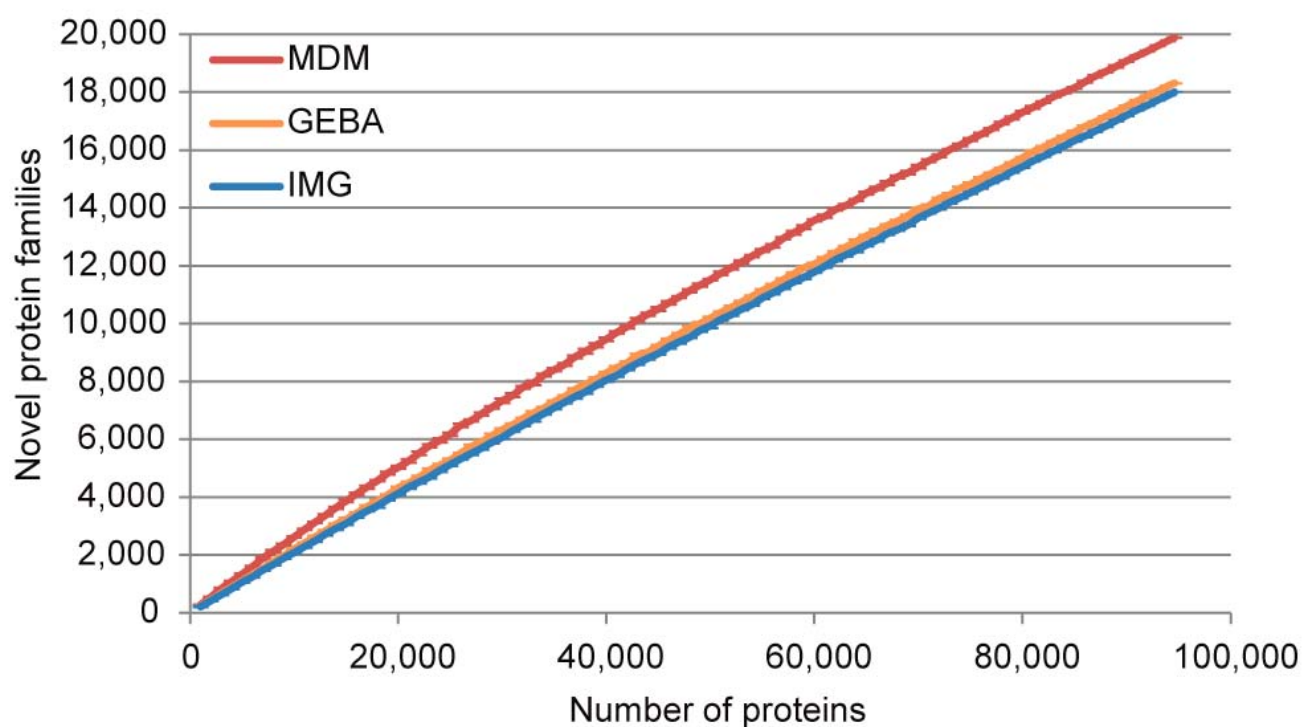
Supplementary Figure 19 | Improved read classification. The circos plot shows the improved read classification by MEGAN4 after the addition of novel single-cell sequenced genomes to the NCBI nr database (right side of dotted line), as compared to the assignments without the MDM SAGs (left side of dotted line). Colored links between the left and right hand-side of the plot indicate the reassignment of about 340 million reads (extrapolated from subsamples of all 475 metagenomes). The thickness of the ribbons on the outer circle indicate the average percent identity of the best BLAST hits for all reads pre- and post-addition of novel SAG data. Note: reads categorized under “no-hits” have by definition no identity value.



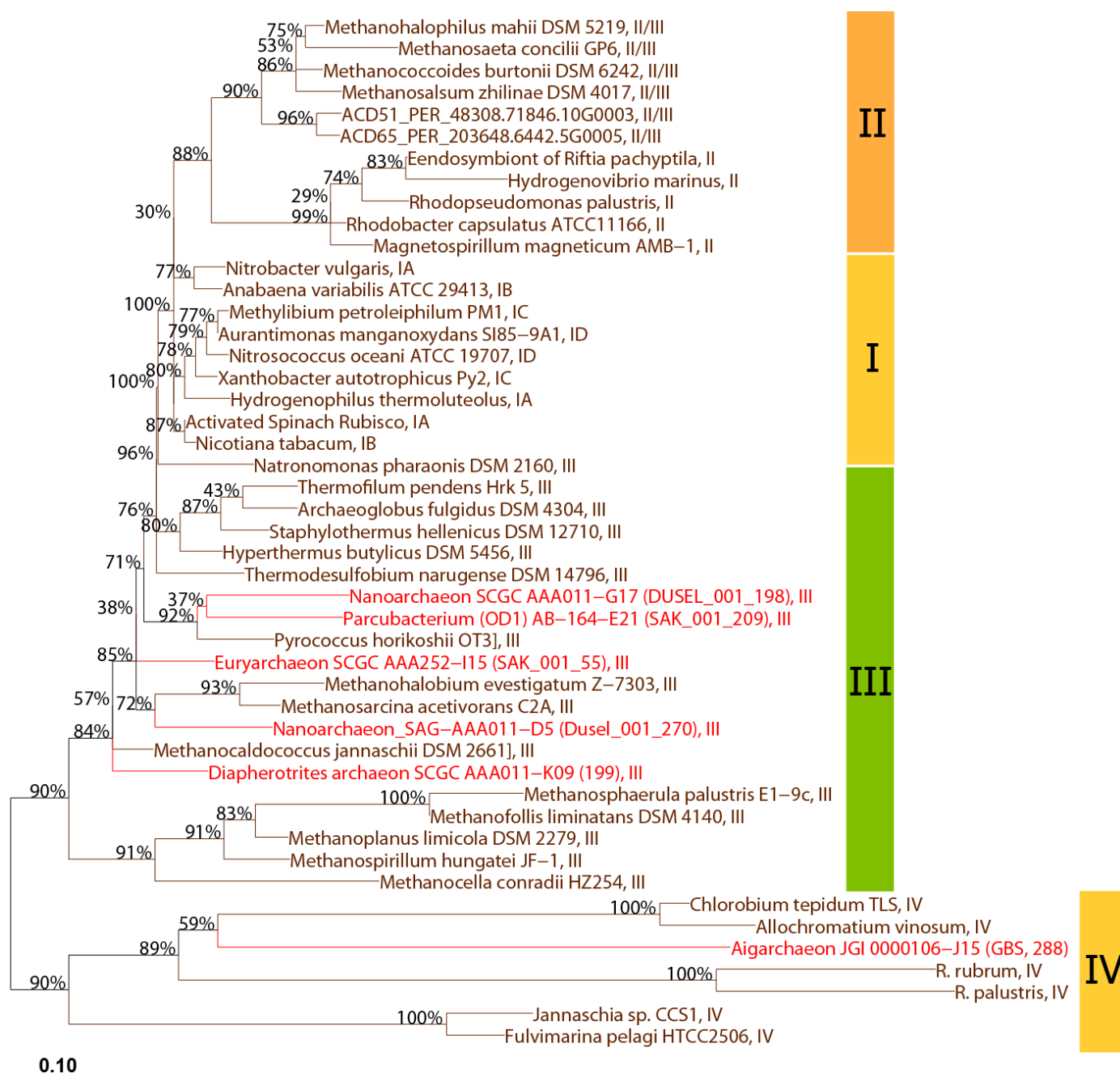
Supplementary Figure 20 | Phylogenetic Diversity (PD) of sequenced and to be sequenced genomes.

(a) Comparison of phylogenetic diversity covered by all sequenced isolates prior to the Genomic Encyclopaedia of Bacteria and Archaea (GEBA) project (Non-GEBA genomes obtained from IMG; blue), with the GEBA genomes added (obtained from IMG; orange), including the Microbial Dark Matter genomes from this study (red), including all microbes classified in the greengenes database without (dark gray) and with (light gray) uncultivated organisms. The PD per genome is 0.24 (Microbial Dark Matter; red), 0.10 (GEBA; orange), and 0.05 (Non-GEBA genomes in the public database; blue). **(b)** Increase of PD of named isolates (cultivated organisms) and clones (uncultivated organisms). PD values were

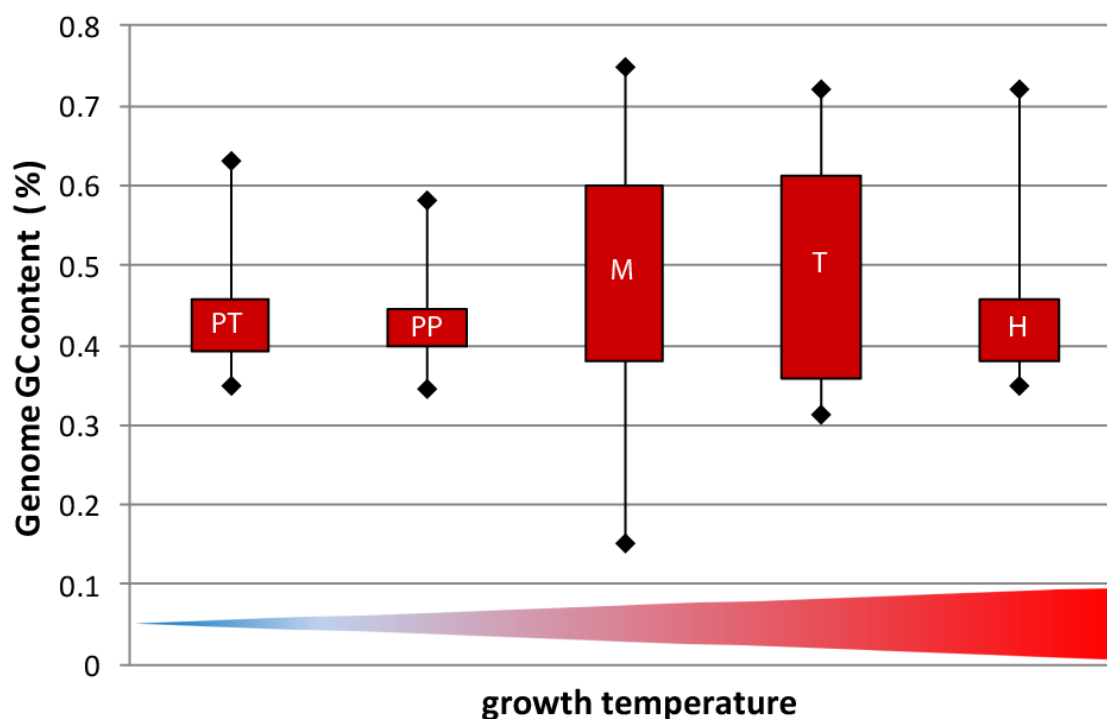
calculated using the sum of the branch length within a SSU rRNA maximum likelihood tree (fasttree, GTR model, CAT approximation).



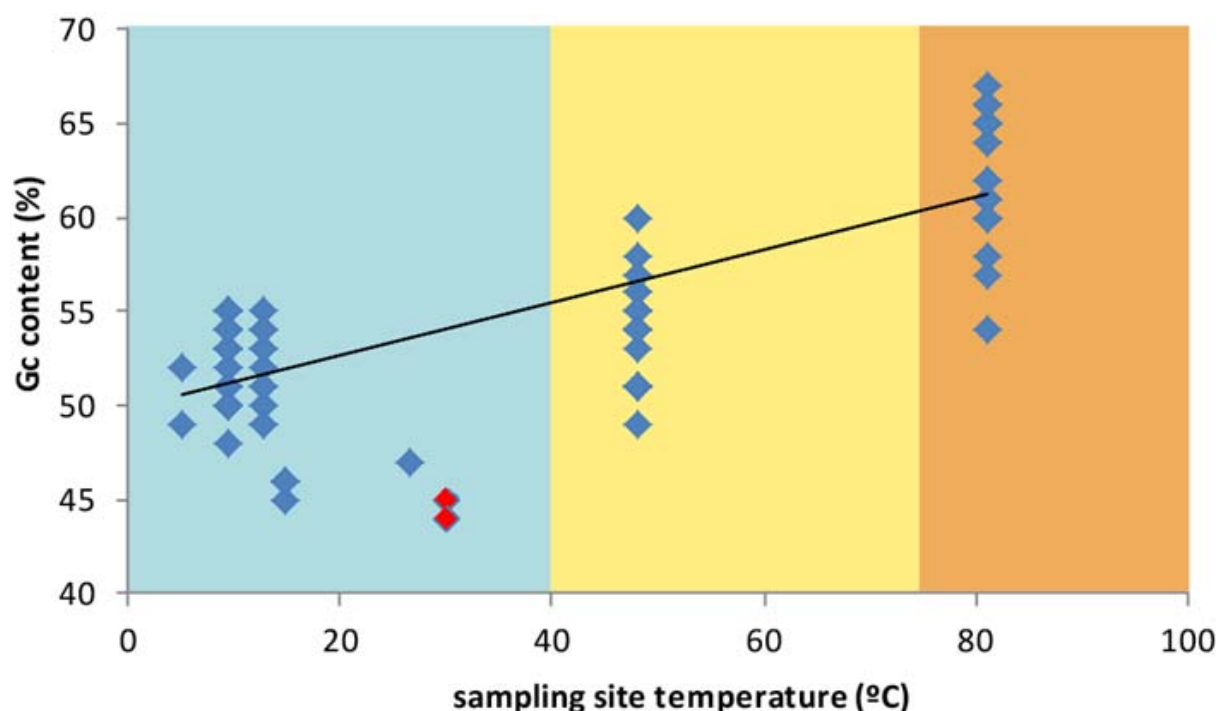
Supplementary Figure 21 | Novel protein families of IMG, GEBA and MDM genomes. Rarefaction curves showing the number of novel protein families (including families with single members) for 100 randomly selected Microbial Dark Matter (MDM), GEBA and isolate genomes from the IMG database. All 94,601 proteins from the MDM dataset were compared to the same number of randomly selected proteins from the GEBA and IMG isolate datasets. Error bars show the standard deviation of 10 repeats per data point. All protein families including singletons are shown (19881).



Supplementary Figure 22 | Phylogeny of putative RuBisCO genes. Phylogenetic maximum likelihood tree (fasttree, JTT, CAT) of Ribulose-1,5-bisphosphate Carboxylase Oxygenase (RuBisCO) genes identified in our SAGs (highlighted in red). The different types of RuBisCO (I to IV) are shown. Four genes from archaeal SAGs and the bacterial candidate phylum Parcubacteria (OD1) could be assigned to RuBisCO type III. The putative RuBisCO gene of the Aigarchaeon SAG (0000106-J15) clusters within type IV (the RuBisCO like protein).



Supplementary Figure 23 | Bacterial genomic GC content and growth temperature range. Genomes are grouped according to their proposed optimal growth temperature from left to right: PT: Psychrotolerant (growth at 7°C or below; n=18), PP: Psychrophile (grow at from −15 °C to +10 °C; n=16), M: Mesophile (20 to 60 °C; n=1205), T: Thermophile (45 to 122 °C; n=99) and H: Hyperthermophile (> 60°C; n=17). Genomic GC content and metadata of all Bacteria were retrieved from IMG (<https://img.jgi.doe.gov/>) by selecting all finished bacterial genomes.

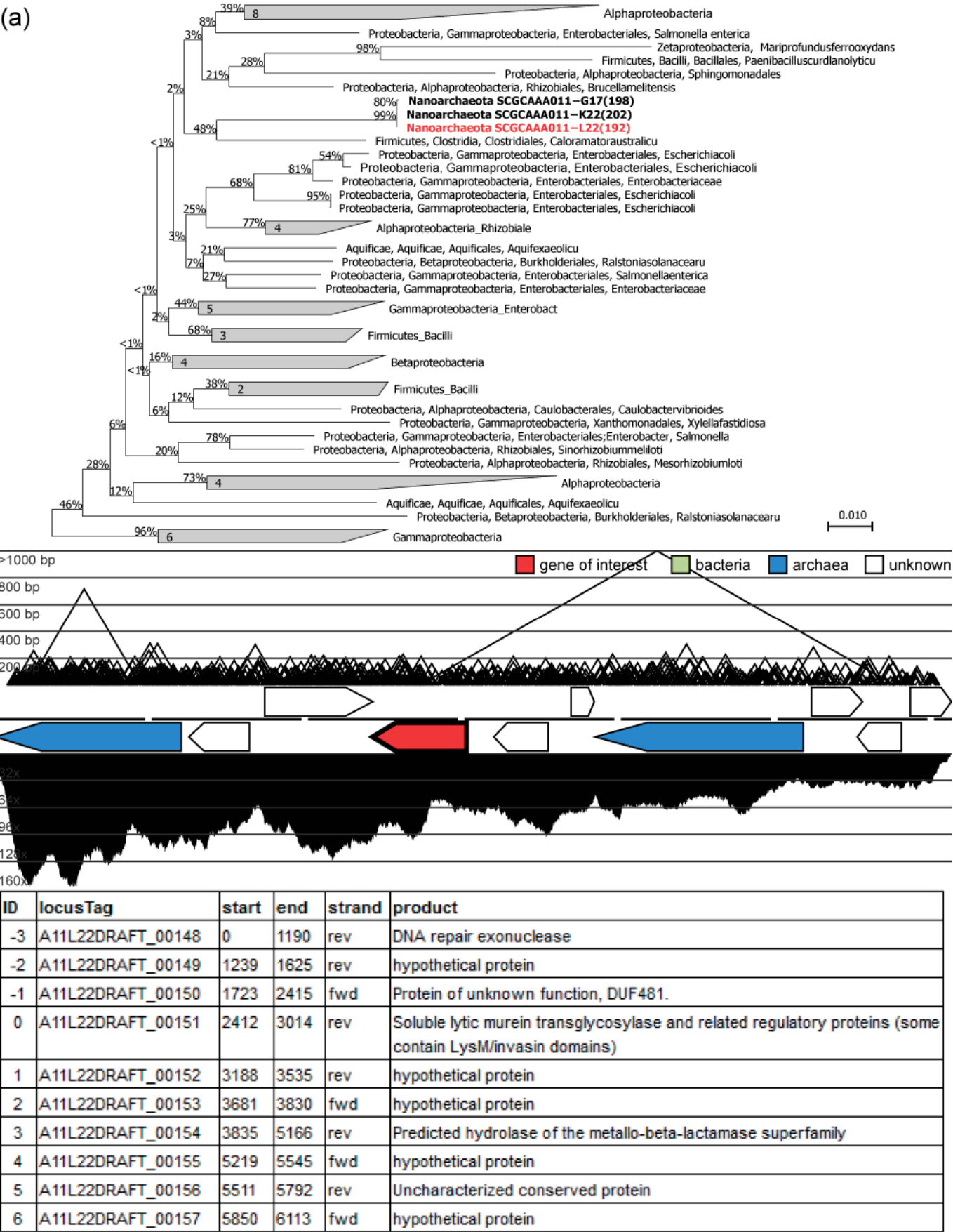


Supplementary Figure 24 | GC content of the SSU rRNA gene and habitat temperatures. Shown is the correlation between habitat temperature and SSU rRNA gene GC content of 175 SAGs for which the sampling site temperature is available. Gracilibacteria (GN02) SAGs from the hydrothermal vent sampling site at the East Pacific Rise (EPR) are outliers with a very low GC content (red symbols). We defined three temperature related groups of organisms based on Wang et al 2006¹: the mesophile group (blue box) contains SAGs sampled at a temperature below 40° C; the thermophile group (yellow box) includes SAGs sampled from 40 to 75 °C; and the hyperthermophile group (orange box) contained SAGs sampled at temperatures greater than 75° C.

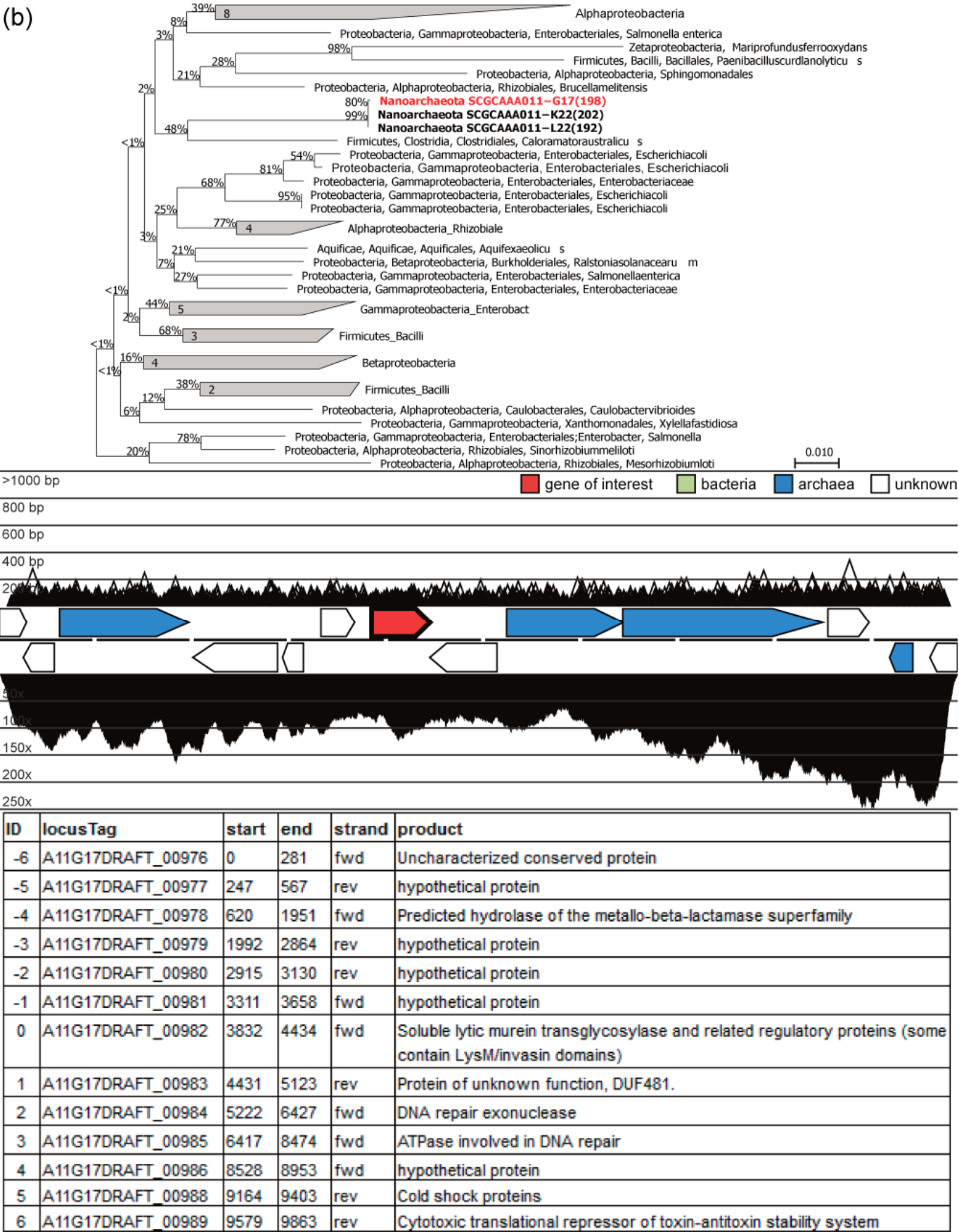
¹ Wang, H.-C., Xia, X. & Hickey, D. Thermal Adaptation of the Small Subunit Ribosomal RNA Gene: A Comparative Study. *Journal of Molecular Evolution* **63**, 120–126 (2006).

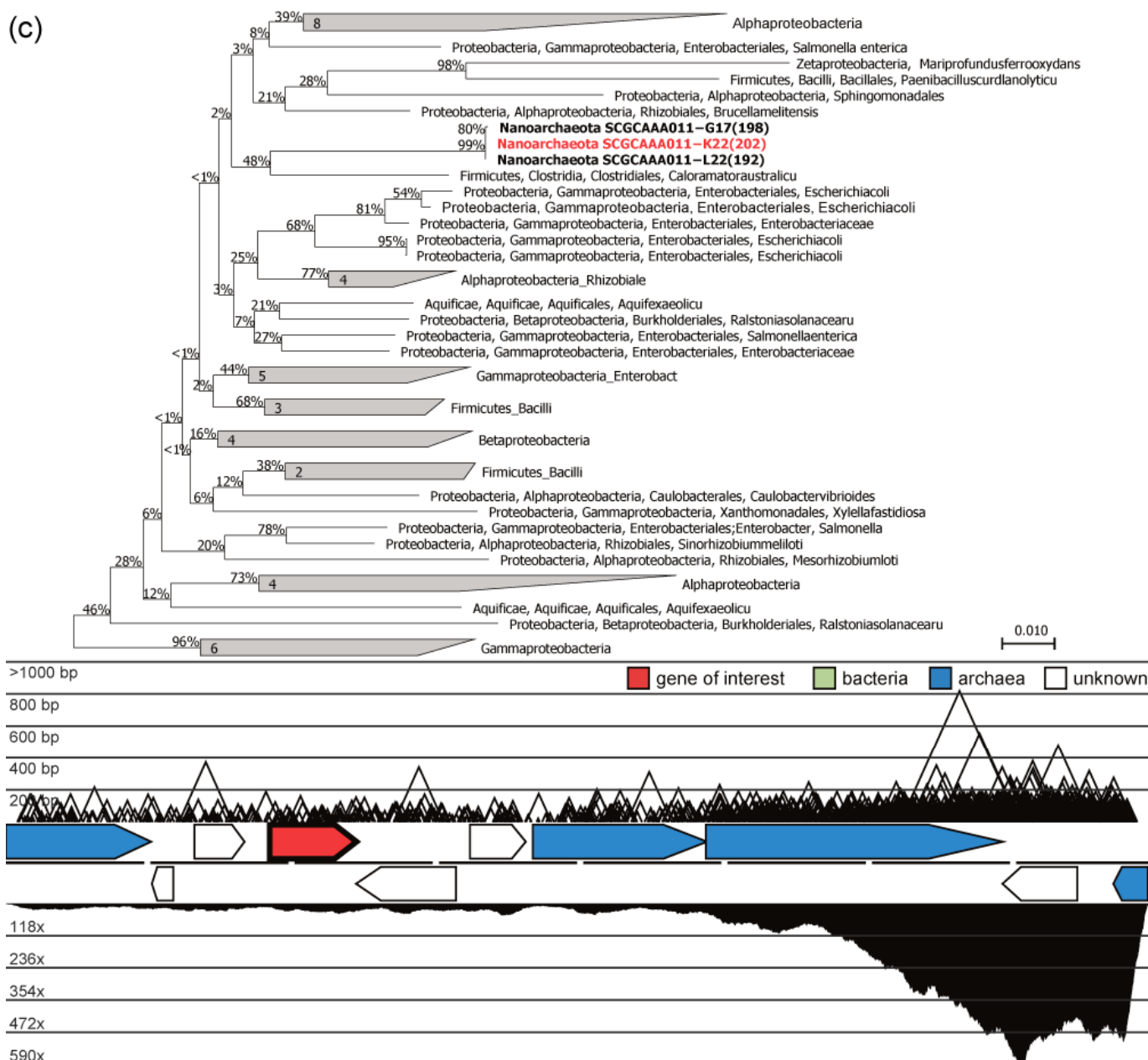
Supplementary Figure 25 | Gene neighborhood analysis. This figure includes all genes, which are involved in the novel findings discussed in the manuscript. The genes are summarized in the table below with the ID referring to the subfigures a-o. Each subfigure focuses on one gene and consists of three parts. The upper part shows the phylogenetic tree highlighting the gene of interest in red. The middle part shows the assembly details of the genes of interest and their genomic neighborhoods. The mapping position of the read pairs are shown above the genes. The height of a link indicates the distance between the two mapped reads (capped at 1000 bp). The central horizontal line symbolizes the genomic location of the gene (in red) with up to 10 kb upstream and downstream regions (in gray). The plot below the genes shows the coverage achieved by mapped reads. The lower part shows details about the gene neighborhood.

ID	Category	Gene product name	Locus tag	DNA length (bp)	Amino acid length (aa)	Genome
a	murein lysis	Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)	A11L22DR AFT_00151	603	200	Nanoarchaeota SCGC AAA011-L22 (192)
b	murein lysis	Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)	A11G17DR AFT_00982	603	200	Nanoarchaeota SCGC AAA011-G17 (198)
c	murein lysis	Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)	A11K22DR AFT_00061	603	200	Nanoarchaeota SCGC AAA011-K22 (202)
d	oxidoreductase	NADH dehydrogenase, FAD-containing subunit	A11D05DR AFT_00286	1050	349	Nanoarchaeota SCGC AAA011-D5 (70)
e	ppGGp _p	(p)ppGpp synthetase, RelA/SpoT family (EC:2.7.6.5)	A11G17DR AFT_00028	1890	629	Nanoarchaeota SCGC AAA011-G17 (198)
f	ppGGp _p	Guanosine polyphosphate pyrophosphohydrolases/synthetases (EC:2.7.6.5)	A11E11DR AFT_00289	1857	618	Diapherotrites SCGC AAA011-E11 (206)
g	ppGGp _p	Guanosine polyphosphate pyrophosphohydrolases/synthetases (EC:2.7.6.5)	A11E11DR AFT_00740	1791	596	Diapherotrites SCGC AAA011-E11 (206)
h	PurP	ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases) (EC:6.3.4.-)	A11A19DR AFT_00045	1014	337	Microgenomates SCGC AAA011-A19 (193)
i	PurP	ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases) (EC:6.3.4.-)	A11L05DR AFT_00194	1014	337	Microgenomates SCGC AAA011-L05 (205)
j	PurP	ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases) (EC:6.3.4.-)	A11L06DR AFT_00258	1017	338	Microgenomates SCGC AAA011-L6 (266)
k	sigma factor	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	A11L22DR AFT_00377	540	179	Nanoarchaeota SCGC AAA011-L22 (192)
l	sigma factor	RNA polymerase sigma factor, sigma-70 family	A11G17DR AFT_01061	540	179	Nanoarchaeota SCGC AAA011-G17 (198)
m	sigma factor	RNA polymerase sigma factor, sigma-70 family	A11K22DR AFT_00364	546	181	Nanoarchaeota SCGC AAA011-K22 (202)
n	sigma factor	RNA polymerase sigma factor, sigma-70 family	A11E11DR AFT_00143	1176	391	Diapherotrites SCGC AAA011-E11 (206)
o	sigma factor	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	A11N19DR AFT_00074	696	231	Diapherotrites SCGC AAA011-N19 (272)

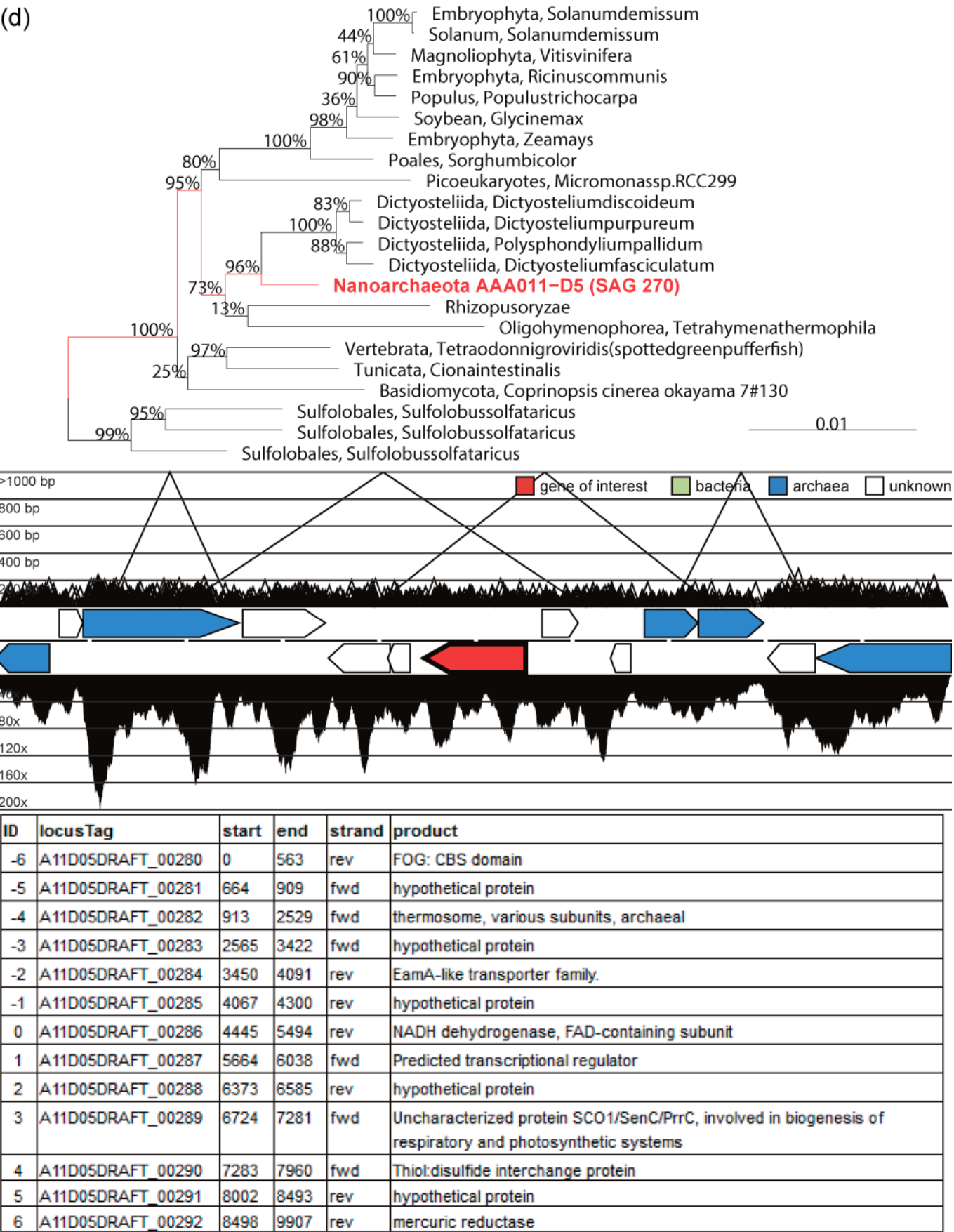


(b)

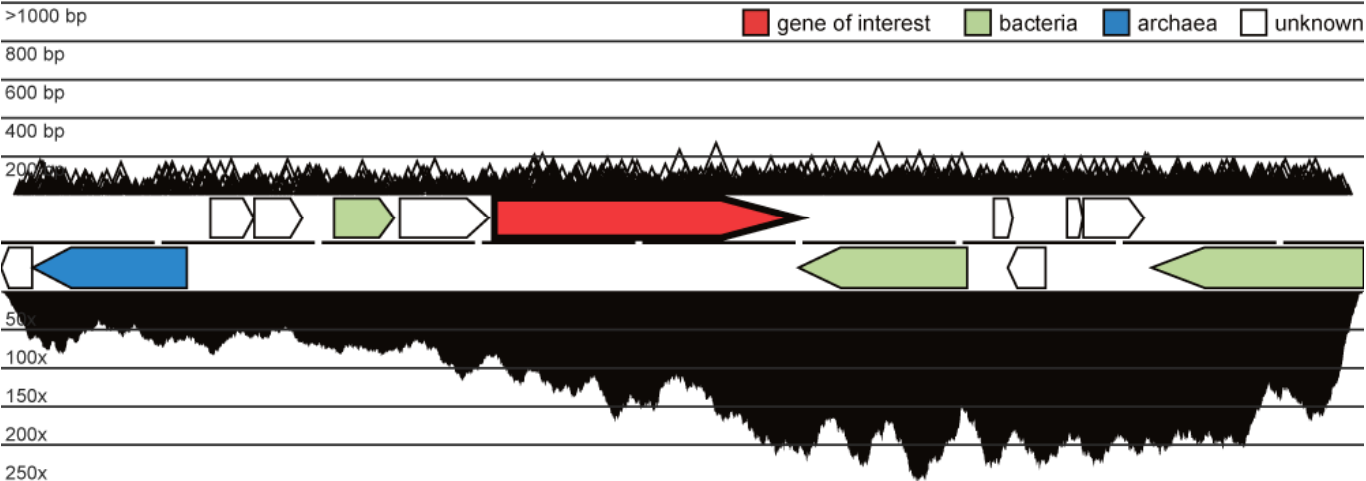
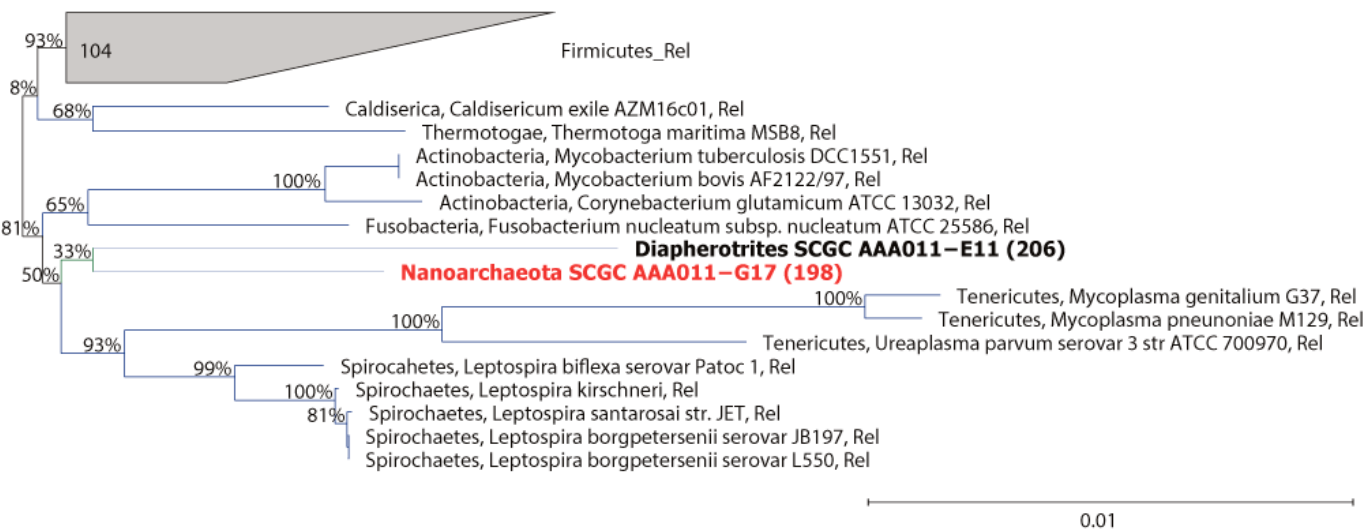




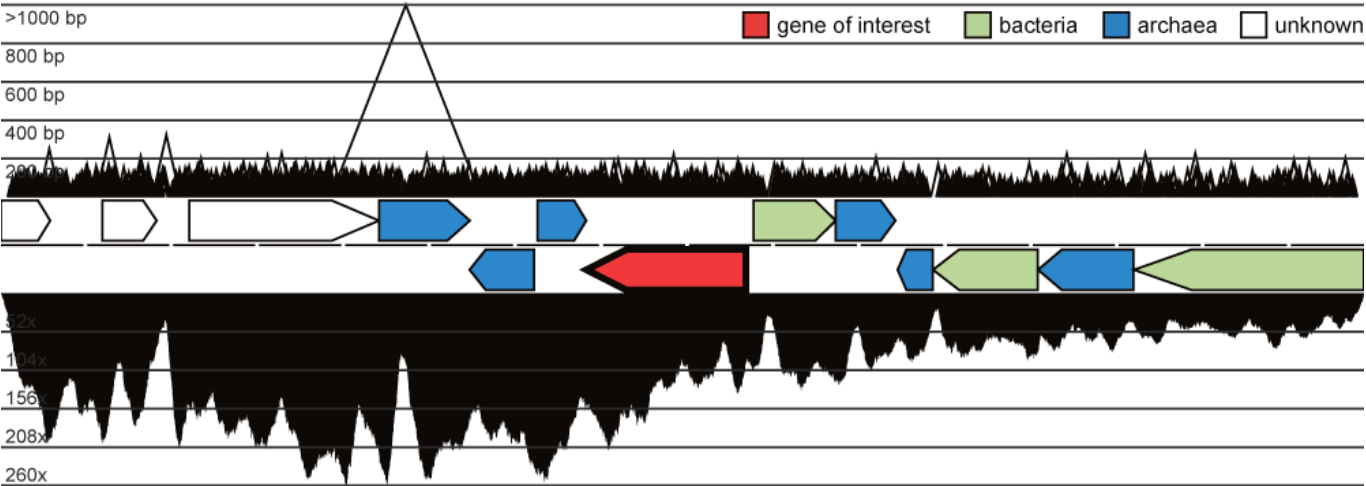
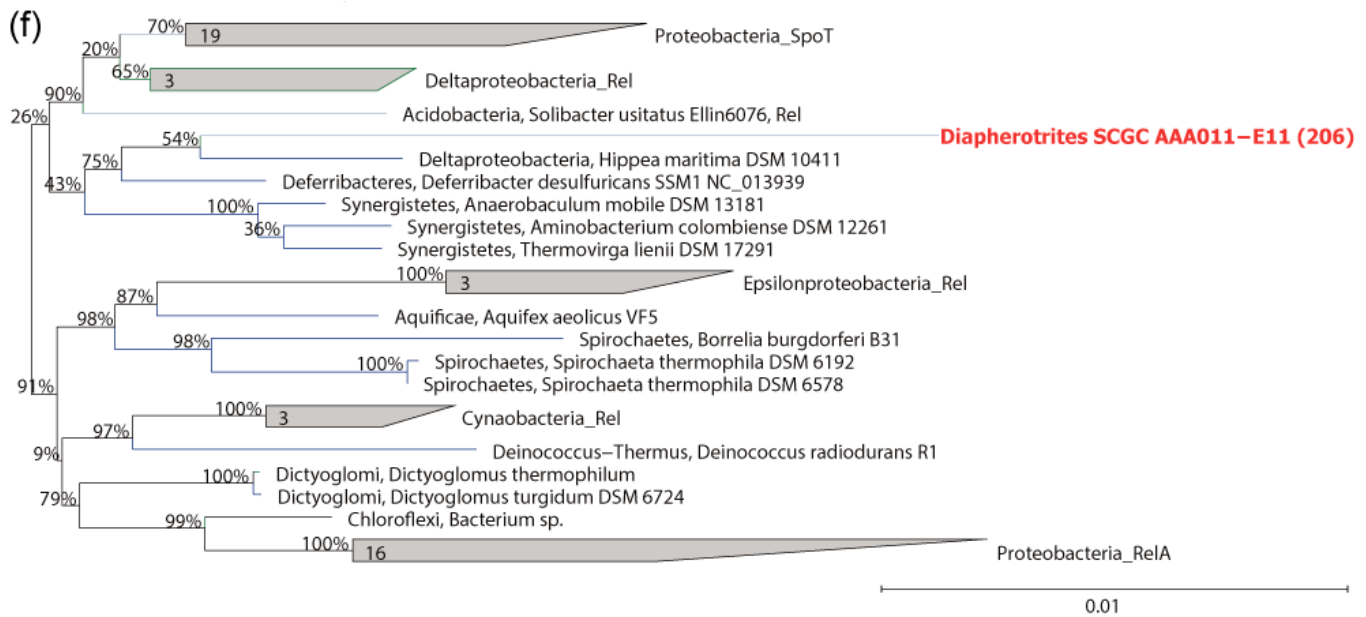
ID	locusTag	start	end	strand	product
-3	A11K22DRAFT_00058	0	1007	fwd	Predicted hydrolase of the metallo-beta-lactamase superfamily
-2	A11K22DRAFT_00059	1012	1161	rev	hypothetical protein
-1	A11K22DRAFT_00060	1307	1654	fwd	hypothetical protein
0	A11K22DRAFT_00061	1828	2430	fwd	Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)
1	A11K22DRAFT_00062	2427	3119	rev	Protein of unknown function, DUF481.
2	A11K22DRAFT_00063	3217	3603	fwd	hypothetical protein
3	A11K22DRAFT_00064	3652	4860	fwd	DNA repair exonuclease
4	A11K22DRAFT_00065	4850	6910	fwd	ATPase involved in DNA repair
5	A11K22DRAFT_00066	6907	7425	rev	hypothetical protein
6	A11K22DRAFT_00068	7675	7914	rev	Cold shock proteins



(e)

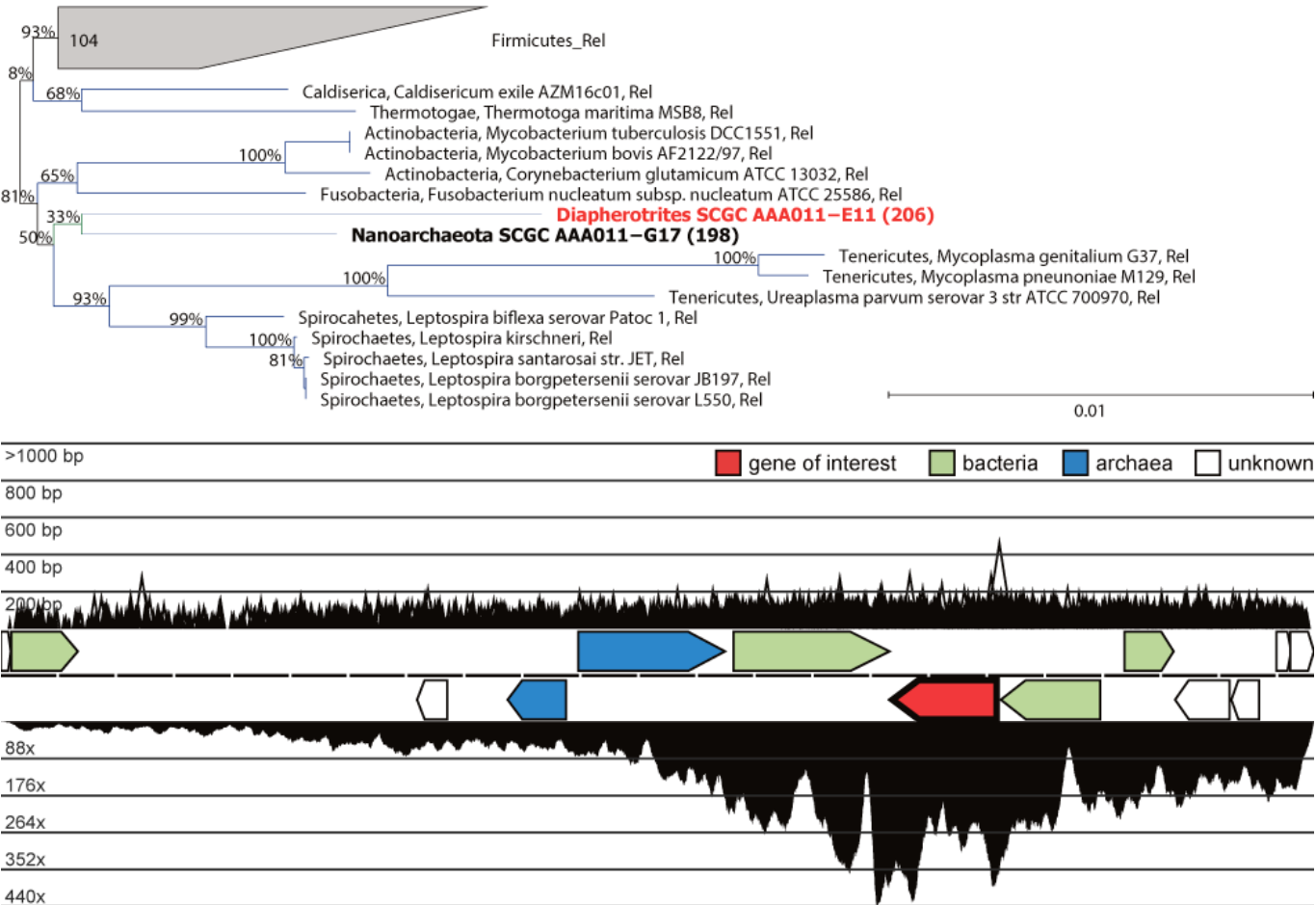


ID	locusTag	start	end	strand	product
-6	A11G17DRAFT_00022	0	194	rev	Sulfur transfer protein involved in thiamine biosynthesis
-5	A11G17DRAFT_00023	197	1159	rev	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2
-4	A11G17DRAFT_00024	1306	1578	fwd	hypothetical protein
-3	A11G17DRAFT_00025	1581	1880	fwd	SpoVT / AbrB like domain.
-2	A11G17DRAFT_00026	2079	2453	fwd	Predicted membrane protein
-1	A11G17DRAFT_00027	2490	3044	fwd	hypothetical protein
0	A11G17DRAFT_00028	3079	4968	fwd	(p)ppGpp synthetase, RelA/SpoT family
1	A11G17DRAFT_00029	4977	6029	rev	hypothetical protein
2	A11G17DRAFT_00030	6195	6314	fwd	hypothetical protein
3	A11G17DRAFT_00031	6283	6519	rev	hypothetical protein
4	A11G17DRAFT_00032	6654	6752	fwd	hypothetical protein
5	A11G17DRAFT_00033	6755	7132	fwd	hypothetical protein
6	A11G17DRAFT_00034	7180	8505	rev	DNA modification methylase



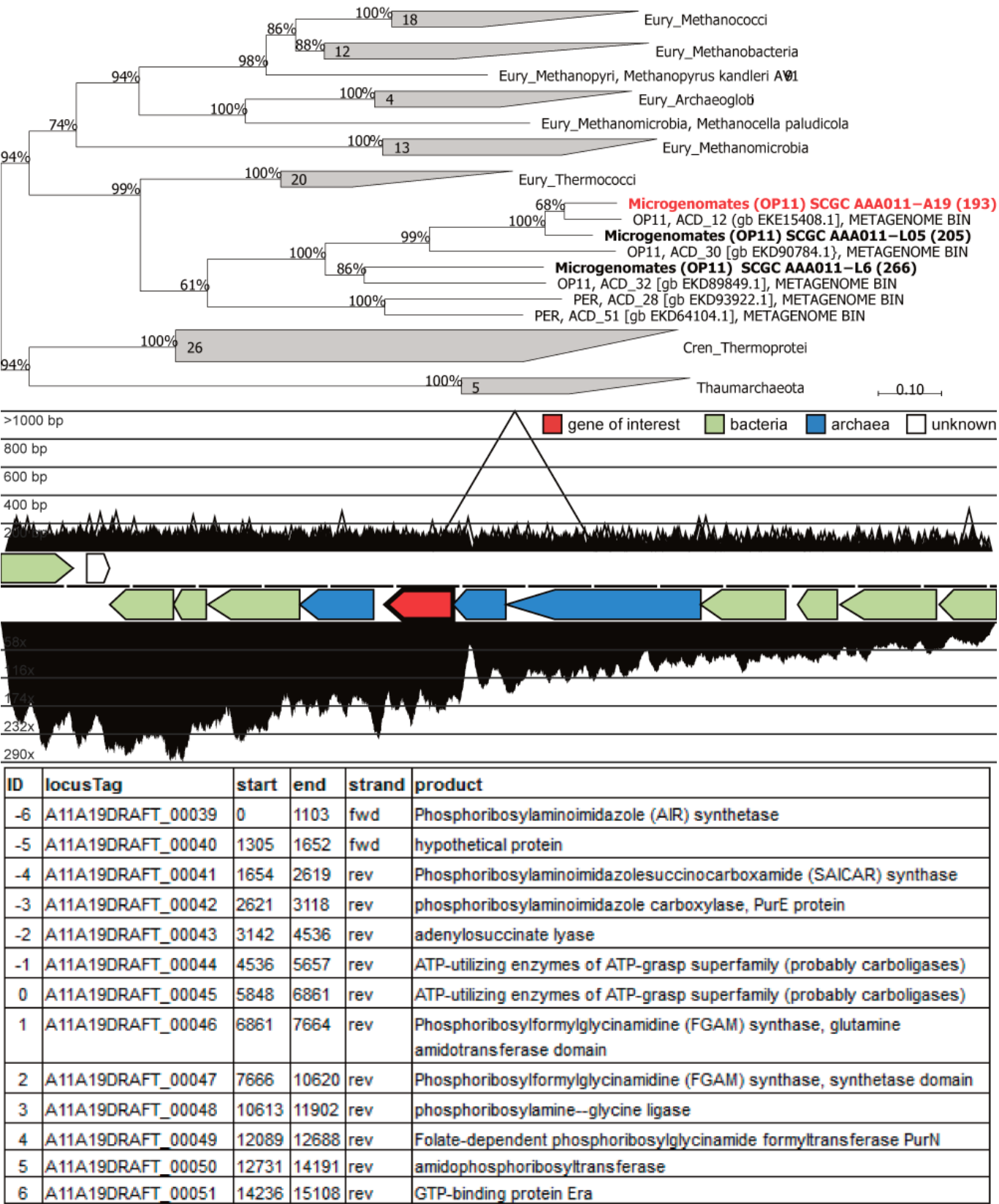
ID	locusTag	start	end	strand	product
-6	A11E11DRAFT_00281	0	572	fwd	hypothetical protein
-5	A11E11DRAFT_00283	1179	1811	fwd	hypothetical protein
-4	A11E11DRAFT_00285	2190	4397	fwd	hypothetical protein
-3	A11E11DRAFT_00286	4399	5454	fwd	Geranylgeranyl pyrophosphate synthase
-2	A11E11DRAFT_00287	5451	6200	rev	RecA-superfamily ATPases implicated in signal transduction
-1	A11E11DRAFT_00288	6242	6808	fwd	ABC-type transport system, involved in lipoprotein release, permease component
0	A11E11DRAFT_00289	6810	8666	rev	Guanosine polyphosphate pyrophosphohydrolases/synthetases
1	A11E11DRAFT_00290	8754	9710	fwd	magnesium Mg(2+) and cobalt Co(2+) transport protein (corA)
2	A11E11DRAFT_00291	9710	10405	fwd	triosephosphate isomerase
3	A11E11DRAFT_00292	10430	10840	rev	SUF system FeS assembly protein, NifU family
4	A11E11DRAFT_00293	10841	12058	rev	cysteine desulfurases, SufS subfamily
5	A11E11DRAFT_00294	12063	13172	rev	ABC-type transport system involved in Fe-S cluster assembly, permease component
6	A11E11DRAFT_00295	13179	15854	rev	FeS assembly protein SufB

(g)

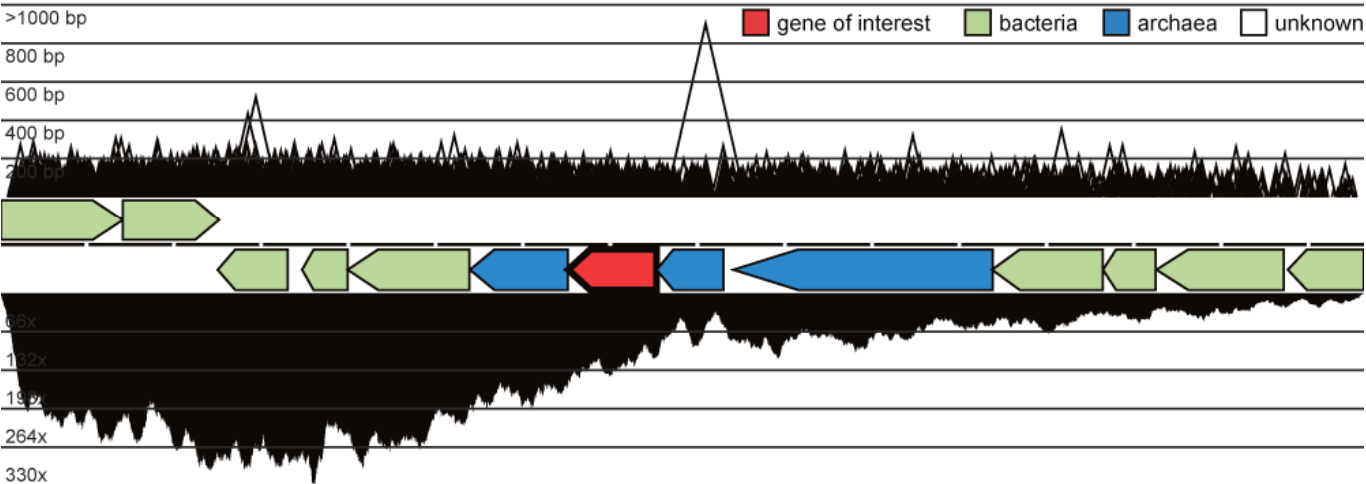
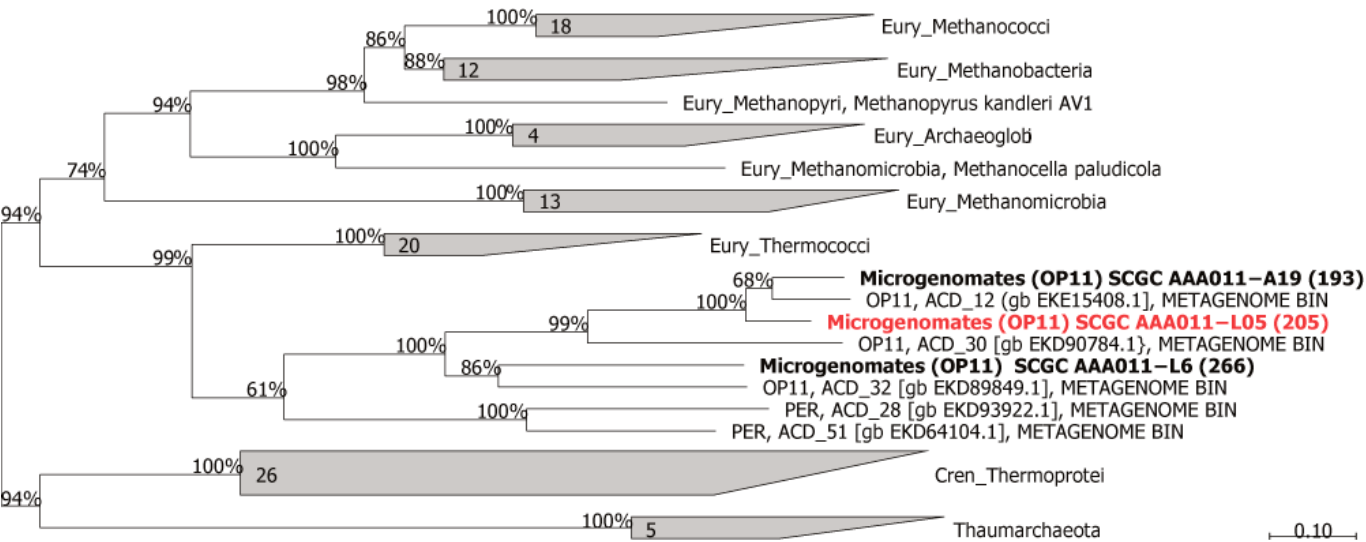


ID	locusTag	start	end	strand	product
-9	A11E11DRAFT_00730	0	167	fwd	hypothetical protein
-8	A11E11DRAFT_00731	180	1325	fwd	Glycosyltransferase
-7	A11E11DRAFT_00732	1748	3058	fwd	16S rRNA. Archaeal SSU
-6	A11E11DRAFT_00733	3430	3503	fwd	tRNA_Ala_TGC
-5	A11E11DRAFT_00734	3871	6826	fwd	23S rRNA. Archaeal LSU
-4	A11E11DRAFT_00735	7168	7686	rev	hypothetical protein
-3	A11E11DRAFT_00737	8728	9732	rev	Aspartate/tyrosine/aromatic aminotransferase
-2	A11E11DRAFT_00738	9942	12473	fwd	alanine--tRNA ligase/alanyl-tRNA synthetase
-1	A11E11DRAFT_00739	12617	15307	fwd	leucyl-tRNA synthetase, eubacterial and mitochondrial family
0	A11E11DRAFT_00740	15358	17148	rev	Guanosine polyphosphate pyrophosphohydrolases/synthetases
1	A11E11DRAFT_00741	17224	18936	rev	arginyl-tRNA synthetase
2	A11E11DRAFT_00743	19352	20200	fwd	Membrane protease subunits, stomatin/prohibitin homologs
3	A11E11DRAFT_00744	20225	21166	rev	hypothetical protein
4	A11E11DRAFT_00745	21195	21674	rev	hypothetical protein
5	A11E11DRAFT_00747	21970	22224	fwd	hypothetical protein
6	A11E11DRAFT_00748	22208	22618	fwd	Predicted nucleotide-binding protein

(h)

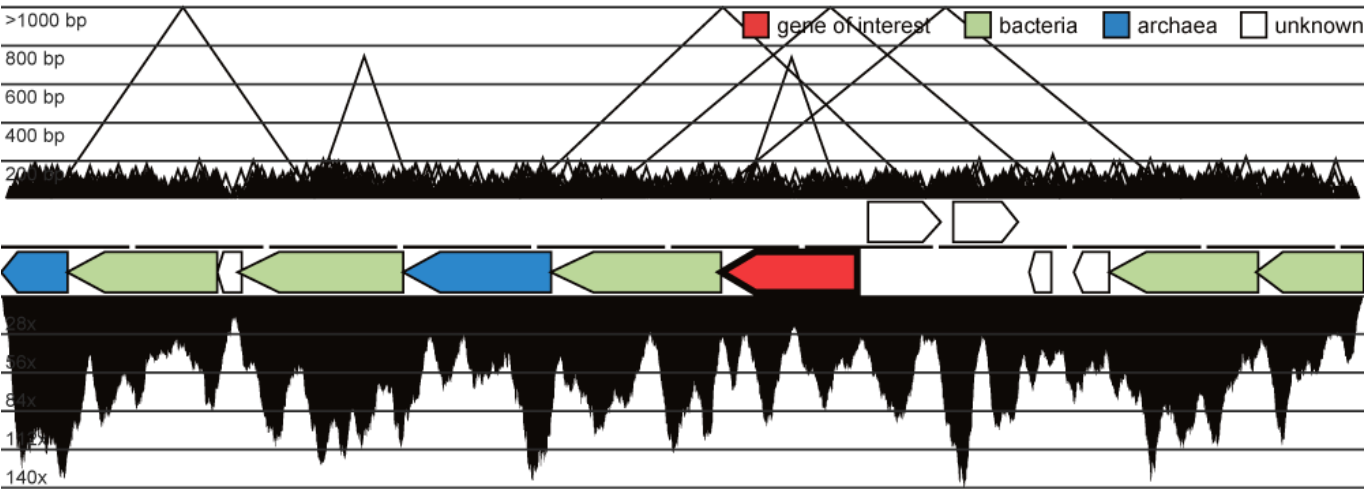
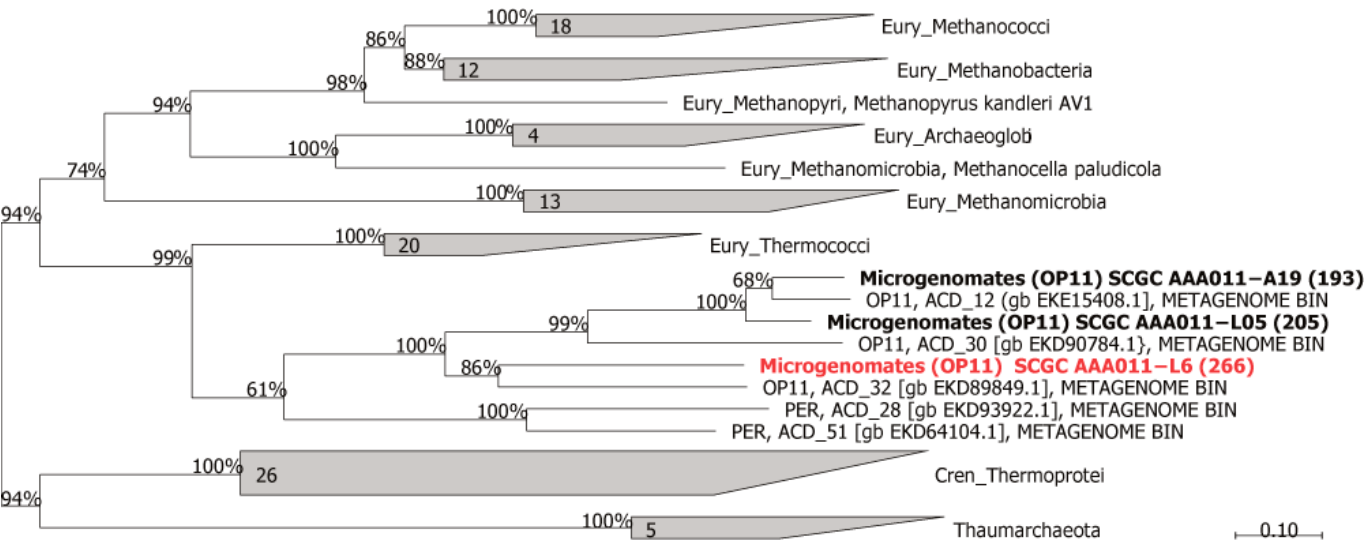


(i)

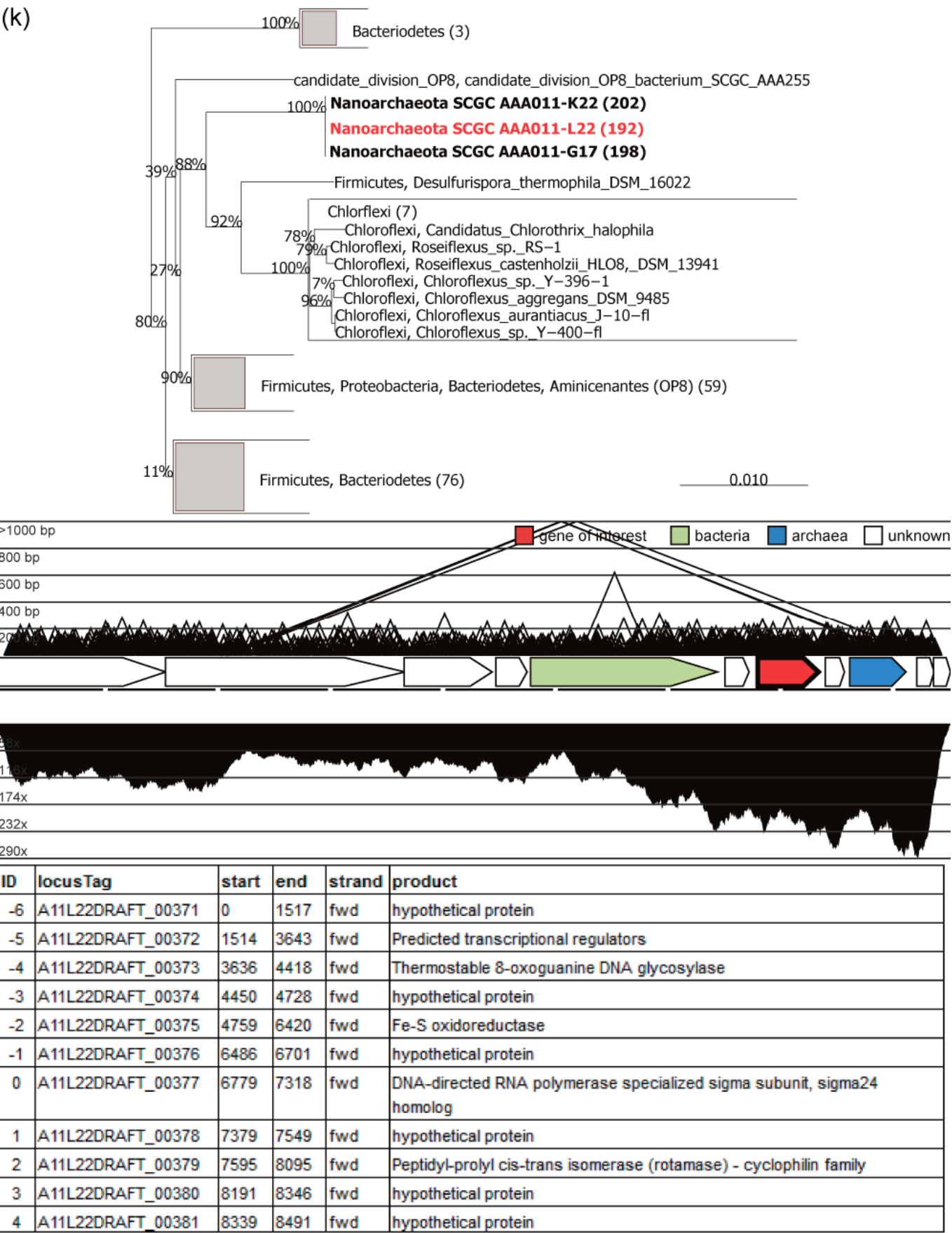


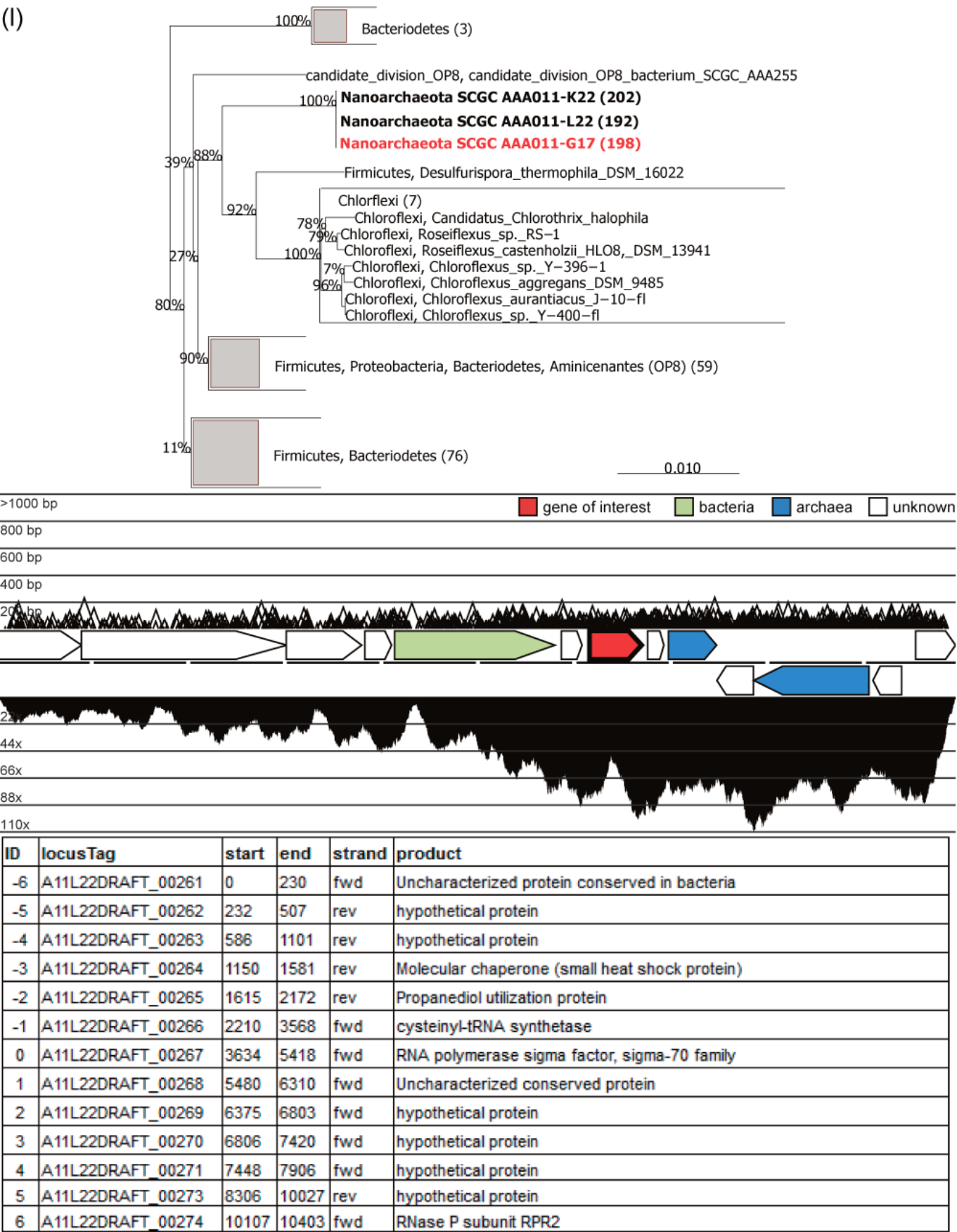
ID	locusTag	start	end	strand	product
-6	A11L05DRAFT_00188	0	1400	fwd	glutamyl-tRNA synthetase, bacterial family
-5	A11L05DRAFT_00189	1397	2500	fwd	Phosphoribosylaminoimidazole (AIR) synthetase
-4	A11L05DRAFT_00190	2483	3286	rev	Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase
-3	A11L05DRAFT_00191	3450	3974	rev	phosphoribosylaminoimidazole carboxylase, PurE protein
-2	A11L05DRAFT_00192	3971	5365	rev	adenylosuccinate lyase
-1	A11L05DRAFT_00193	5372	6493	rev	ATP-utilizing enzymes of ATP-grasp superfamily (probably carbogligases)
0	A11L05DRAFT_00194	6498	7511	rev	ATP-utilizing enzymes of ATP-grasp superfamily (probably carbogligases)
1	A11L05DRAFT_00195	7511	8278	rev	Phosphoribosylformylglycinamide (FGAM) synthase, glutamine amidotransferase domain
2	A11L05DRAFT_00196	8390	11362	rev	Phosphoribosylformylglycinamide (FGAM) synthase, synthetase domain
3	A11L05DRAFT_00197	11355	12623	rev	phosphoribosylamine--glycine ligase
4	A11L05DRAFT_00198	12625	13227	rev	Folate-dependent phosphoribosylglycinamide formyltransferase PurN
5	A11L05DRAFT_00199	13237	14697	rev	amidophosphoribosyltransferase
6	A11L05DRAFT_00200	14743	15615	rev	GTP-binding protein Era

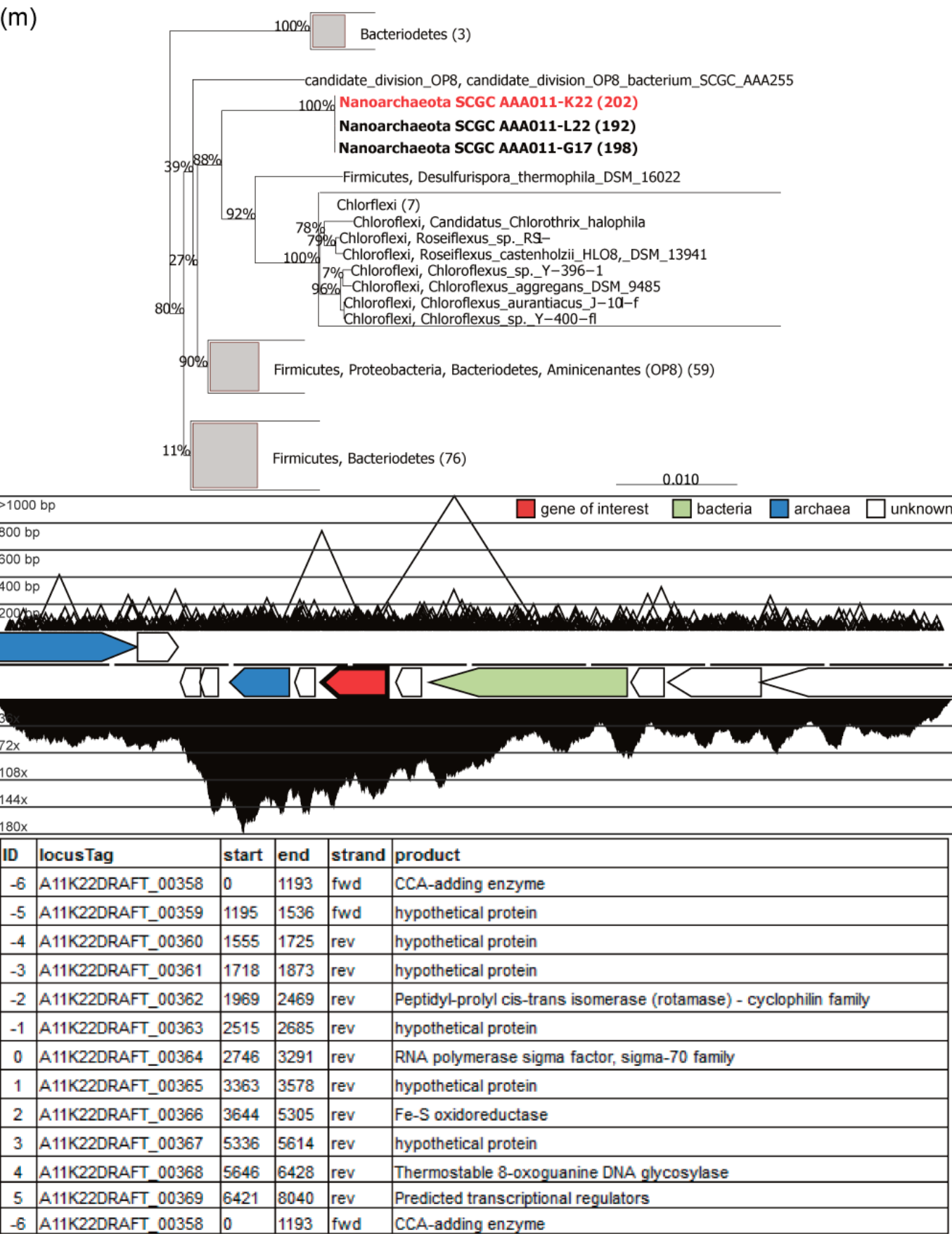
(j)



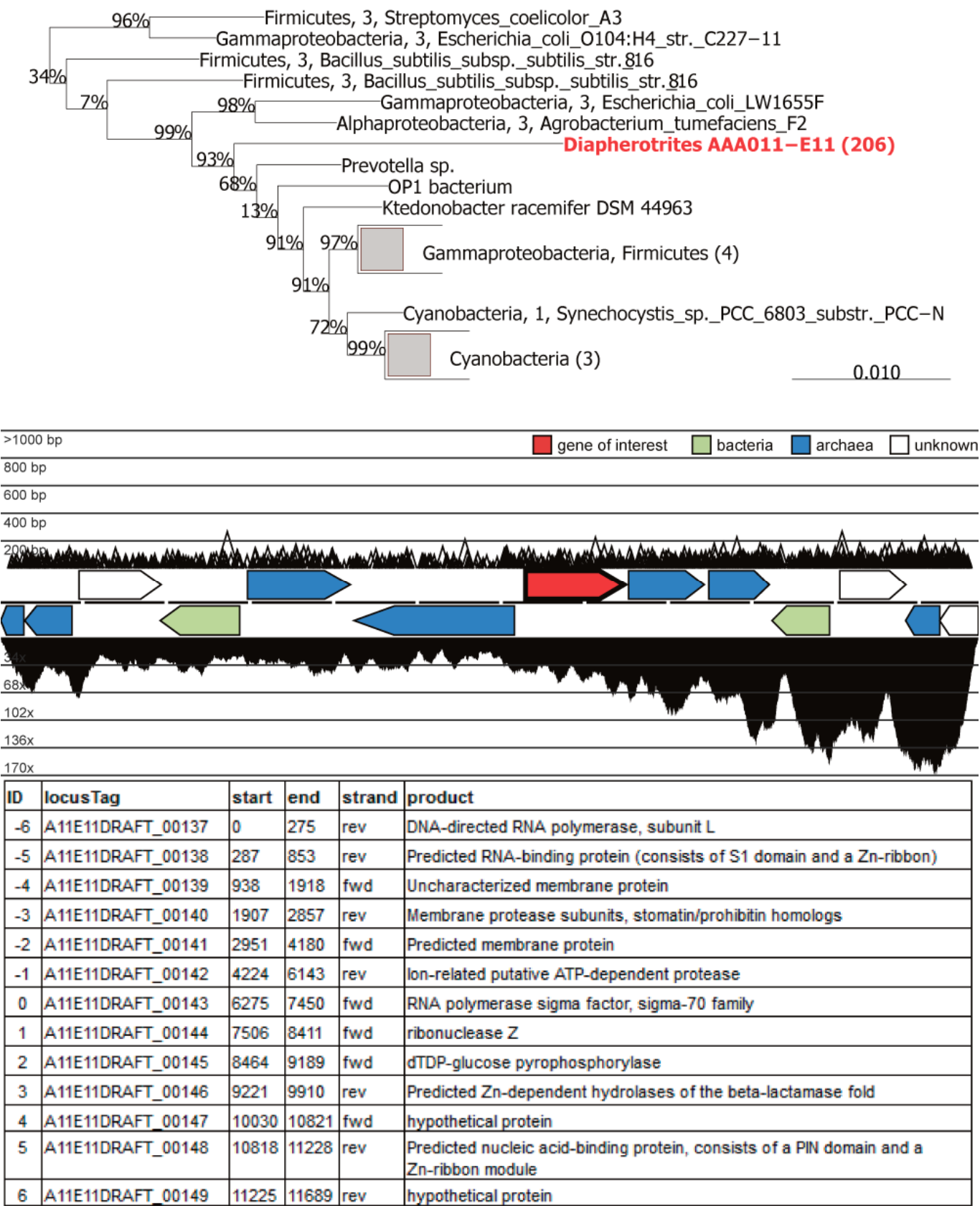
ID	locusTag	start	end	strand	product
-6	A11L06DRAFT_00252	0	497	rev	phosphoribosylaminoimidazole carboxylase, PurE protein
-5	A11L06DRAFT_00253	494	1615	rev	phosphoribosylaminoimidazole carboxylase, PurK protein
-4	A11L06DRAFT_00254	1615	1797	rev	Adenylosuccinate lyase C-terminus.
-3	A11L06DRAFT_00255	1766	3004	rev	adenylosuccinate lyase
-2	A11L06DRAFT_00256	3001	4107	rev	ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases)
-1	A11L06DRAFT_00257	4104	5378	rev	adenylosuccinate synthase
0	A11L06DRAFT_00258	5378	6394	rev	ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases)
1	A11L06DRAFT_00259	6474	7016	fwd	Acetyltransferases
2	A11L06DRAFT_00260	7108	7593	fwd	hypothetical protein
3	A11L06DRAFT_00261	7677	7841	rev	hypothetical protein
4	A11L06DRAFT_00262	8010	8273	rev	hypothetical protein
5	A11L06DRAFT_00263	8273	9385	rev	Phosphoribosylaminoimidazole (AIR) synthetase
6	A11L06DRAFT_00264	9369	10175	rev	phosphoribosylformylglycinamide synthase I



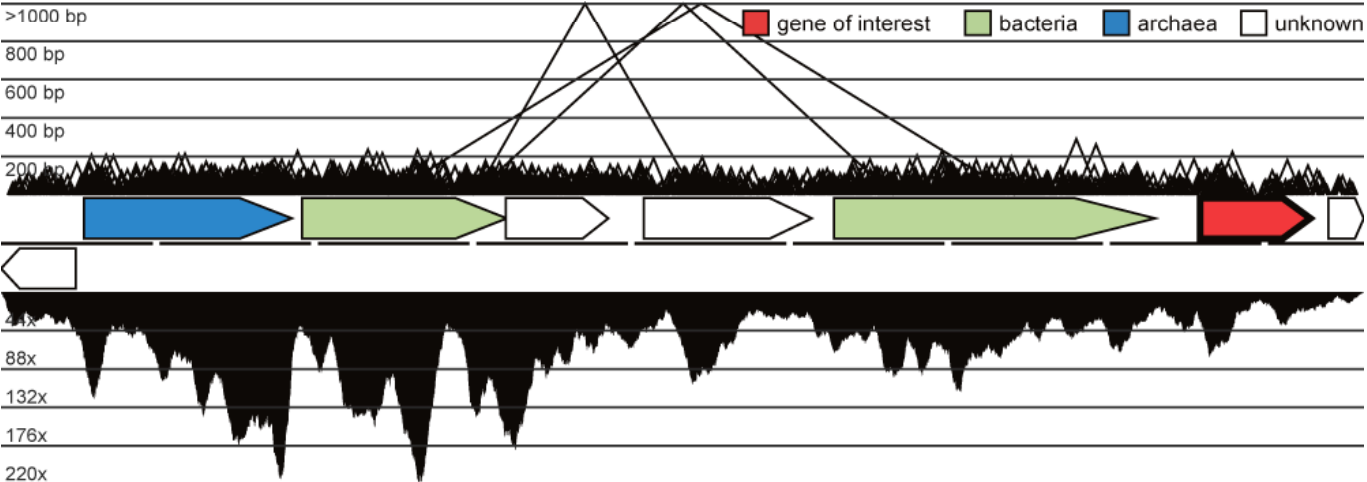
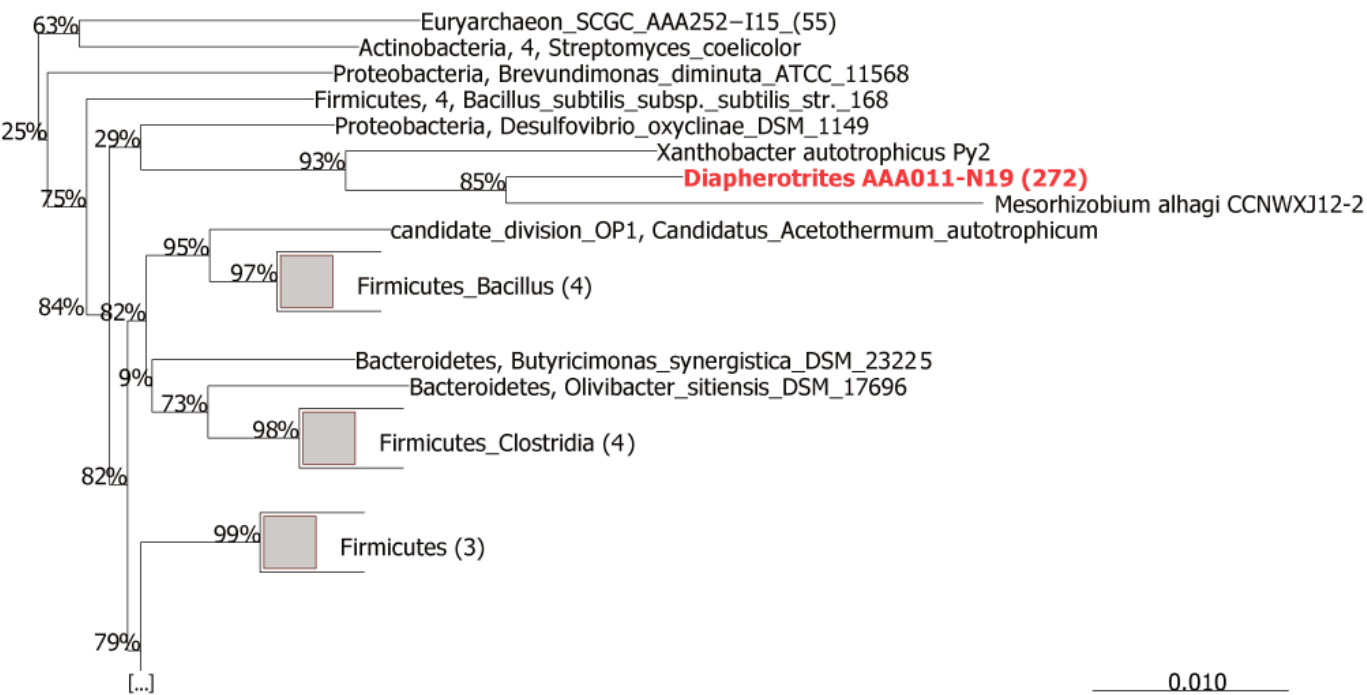




(n)



(o)



ID	locusTag	start	end	strand	product
-6	A11N19DRAFT_00066	0	473	rev	hypothetical protein
-5	A11N19DRAFT_00067	524	1837	fwd	Predicted ATPase
-4	A11N19DRAFT_00068	1902	3194	fwd	Glycosyltransferases, probably involved in cell wall biogenesis
-3	A11N19DRAFT_00069	3188	3835	fwd	hypothetical protein
-2	A11N19DRAFT_00071	4059	5120	fwd	hypothetical protein
-1	A11N19DRAFT_00072	5261	7288	fwd	hypothetical protein
0	A11N19DRAFT_00074	7573	8268	fwd	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
1	A11N19DRAFT_00075	8382	8606	fwd	hypothetical protein

Supplementary Tables

Supplementary Table 1 | Genome sequences obtained from uncultivated microorganisms. The method used to obtain the genomes is shown as (meta) metagenome assembly bin(s) and (single) single cell genomics, or both. Alternative taxonomic names are provided in parenthesis and novel phyla names proposed in this study are shown in square brackets (see also Supplementary Table 5).

taxonomic group	candidatus organism(s)	method	reference
Aigarchaeota	' <i>Caldiarchaeum subterraneum</i> '	meta	Nunoura, T. <i>et al.</i> 2011 ¹
ARMAN [Parvarchaeota]	' <i>Micrarchaeum acidiphilum</i> ARMAN-2', ' <i>Parvarchaeum acidiphilum</i> ARMAN-4', ' <i>Parvarchaeum acidophilus</i> ARMAN-5'	meta	Baker, B. J. <i>et al.</i> 2010 ²
GNO2 (BD1-5) [Gracilibacteria]	5 unnamed draft genomes	meta	Wrighton, K. C. <i>et al.</i> 2012 ³
Korarchaeota	' <i>Korarchaeum cryptophylum</i> '	meta	Elkins, J. G. <i>et al.</i> 2008 ⁴
Nanohaloarchaea [Nanohaloarchaeota]	<i>Candidatus Haloredivivus</i> <i>Candidatus Nanosalina</i> , <i>Candidatus</i> <i>Nanosalarum</i>	single meta	Ghai, R. <i>et al.</i> 2011 ⁵ Narasingarao, P. <i>et al.</i> 2012 ⁶
NC10	' <i>Methyloirabilis oxyfera</i> '	meta	Ettwig, K. F. <i>et al.</i> 2010 ⁷
OD1 [Parcubacteria]	18 unnamed draft genomes	meta	Wrighton, K. C. <i>et al.</i> 2012 ³
OP1 [Acetothermia]	' <i>Acetothermum autotrophicum</i> '	meta	Takami, H. <i>et al.</i> 2010 ⁸
OP11 [Microgenomates]	18 unnamed draft genomes	meta	Wrighton, K. C. <i>et al.</i> 2012 ³
OP11 [Microgenomates]	ZG1 draft genome	single	Youssef, N. H., <i>et al.</i> 2011 ⁹
OP9 [Atribacteria]	Two draft genomes, OP9-cSCG and OP9-77CS	meta + single	Dodsworth, J. <i>et al.</i> 2013 ¹⁰
Poribacteria	One draft genome, WGA-A3	single	Siegl, A. <i>et al.</i> 2011 ¹¹
SR1	SR1-OR1 genome	meta + single	Campbell <i>et al.</i> 2013 ¹²
TM7	5 draft genomes combined to TM7_GTL1	single	Podar, M. <i>et al.</i> 2007 ¹³
TM7 (Saccharobacteria)	' <i>Saccharobacterium alaburgensis</i> ' + three other TM7 genomes	meta	Albertsen, M. <i>et al.</i> 2013 ¹⁴
WWE1 [Cloacimonetes]	' <i>Cloacamonas acidaminovorans</i> ' [sic]	meta	Pelletier, E. <i>et al.</i> 2008 ¹⁵

¹ Nunoura, T. *et al.* Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* **39**, 3204–3223 (2011).

² Baker, B. J. *et al.* Enigmatic, ultrasmall, uncultivated Archaea. *PNAS* **107**, 8806–8811 (2010).

³ Wrighton, K. C. *et al.* Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661–1665 (2012).

⁴ Elkins, J. G. *et al.* A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences* **105**, 8102–8107 (2008).

⁵ Ghai, R. *et al.* New Abundant Microbial Groups in Aquatic Hypersaline Environments. *Sci. Rep.* **1**, (2011).

⁶ Narasingarao, P. *et al.* De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J* **6**, 81–93 (2012).

⁷ Ettwig, K. F. *et al.* Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**, 543–548 (2010).

⁸ Takami, H. *et al.* A Deeply Branching Thermophilic Bacterium with an Ancient Acetyl-CoA Pathway Dominates a Subsurface Ecosystem. *PLoS ONE* **7**, e30559 (2012).

⁹ Youssef, N. H. *et al.* A partial genome assembly from a candidate division OP11 single cell from an anoxic spring (Zodlone spring, OK, USA). *Appl. Environ. Microbiol.* AEM.06059–11 (2011). doi:10.1128/AEM.06059-11

¹⁰ Dodsworth, J. *et al.* Single-cell and metagenomic analyses indicate a fermentative, saccharolytic lifestyle for members of the OP9 lineage. *Nature Communications*. **In press**.

¹¹ Siegl, A. *et al.* Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME J* **5**, 61–70 (2011).

¹² Campbell, J. H. *et al.* UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *PNAS* **110**, 5540–5545 (2013).

¹³ Podar, M. *et al.* Targeted Access to the Genomes of Low-Abundance Organisms in Complex Microbial Communities. *Appl. Environ. Microbiol.* **73**, 3205–3214 (2007).

¹⁴ Albertsen M. *et. al.* Recovery of genomes from rare, uncultured bacteria by differential coverage binning of multiple deep metagenomes. *Nature Biotechnology*, accepted (2013).

¹⁵ Pelletier, E. *et al.* ‘Candidatus Cloacamonas Acidaminovorans’: Genome Sequence Reconstruction Provides a First Glimpse of a New Bacterial Division. *J. Bacteriol.* **190**, 2572–2579 (2008).

Supplementary Table 2 | Genome statistics of all 201 SAGs and 21 combined assemblies. Shown are the Microbial Dark Matter ID (MDM ID), taxonomic group, name of the SAG, sampling site (site), assembly size in bp (assi), GC content (GC), number of contigs (#con), contig N50 in bp (N50), size of the largest contig in bp (lcon), coding density as percentage of coding sequence (CDS), the estimated completeness of the single-cell genomes (comp), and the operational taxonomic unit (OUT) number for the combined assemblies.

SAG assemblies

MDM ID	taxonomic group	SAG name	site	assi	GC	#con	N50	lcon	CDS	comp
122	Acetothermia	AAA255-C06	SAK	669384	55%	85	9935	33483	86%	18%
186	Aenigmarchaeota	AAA011-F07	HSM	248182	53%	18	28977	44983	85%	23%
271	Aenigmarchaeota	AAA011-O16	HSM	428886	39%	15	41863	57613	95%	47%
281	Aenigmarchaeota	0000106-F11	GBS	361035	37%	24	35139	70790	90%	36%
70	Aerophobetes	AAA252-L23	SAK	2305959	43%	190	19503	58784	86%	67%
129	Aerophobetes	AAA255-F10	SAK	1263620	40%	102	19720	59665	86%	36%
222	Aerophobetes	0000014-C22	ETL	428566	39%	51	11003	26827	83%	16%
223	Aerophobetes	0000014-A15	ETL	368019	37%	32	17229	42484	83%	14%
1	Aigarchaeota	AAA471-A16	GBS	253217	58%	23	14393	35496	90%	12%
4	Aigarchaeota	AAA471-B22	GBS	267561	62%	17	40278	65831	91%	18%
9	Aigarchaeota	AAA471-D15	GBS	507023	62%	27	37213	70829	91%	44%
10	Aigarchaeota	AAA471-E14	GBS	482646	55%	27	19910	49102	93%	43%
11	Aigarchaeota	AAA471-E16	GBS	525966	61%	39	24529	55621	91%	33%
13	Aigarchaeota	AAA471-F17	GBS	741841	62%	43	26075	109193	91%	64%
14	Aigarchaeota	AAA471-G05	GBS	640245	62%	39	31581	75493	91%	39%
17	Aigarchaeota	AAA471-I13	GBS	164795	61%	18	11690	46288	90%	14%
18	Aigarchaeota	AAA471-J07	GBS	249792	62%	19	18425	33878	92%	17%
19	Aigarchaeota	AAA471-J08	GBS	252228	61%	16	26696	37595	93%	17%
112	Aigarchaeota	0000001-A7	GBS	412511	54%	33	16590	36835	93%	45%
113	Aigarchaeota	0000001-B8	GBS	373687	62%	36	15717	32201	91%	23%
118	Aigarchaeota	0000001-H6	GBS	198883	59%	21	13490	22540	88%	10%
288	Aigarchaeota	0000106-J15	GBS	1085321	38%	34	56092	103790	87%	68%
31	Aminicenantes	AAA252-A02	SAK	1140667	39%	81	23816	65759	89%	25%
33	Aminicenantes	AAA252-B17	SAK	965122	39%	74	20371	79923	90%	34%
36	Aminicenantes	AAA252-C08	SAK	1087559	39%	81	24394	90155	89%	42%
38	Aminicenantes	AAA252-D06	SAK	1851771	39%	98	35278	87646	89%	53%
40	Aminicenantes	AAA252-D18	SAK	1408933	39%	99	24357	84295	89%	52%
46	Aminicenantes	AAA252-F08	SAK	1125211	39%	104	17780	61885	88%	56%
47	Aminicenantes	AAA252-G04	SAK	978461	39%	102	13506	55239	88%	31%
48	Aminicenantes	AAA252-G05	SAK	1132687	39%	78	23888	65603	89%	26%
49	Aminicenantes	AAA252-G06	SAK	1473924	39%	98	25407	67935	88%	42%
52	Aminicenantes	AAA252-G21	SAK	1452805	39%	95	30774	104029	89%	45%
54	Aminicenantes	AAA252-I14	SAK	1469200	39%	103	24364	51491	88%	46%
57	Aminicenantes	AAA252-J09	SAK	1920737	39%	107	33472	117648	89%	62%
58	Aminicenantes	AAA252-J15	SAK	1005574	39%	71	23068	70591	89%	18%
59	Aminicenantes	AAA252-J21	SAK	1361089	39%	87	29978	89686	89%	43%
60	Aminicenantes	AAA252-K06	SAK	1184457	41%	102	17951	99460	90%	55%
61	Aminicenantes	AAA252-K07	SAK	1830549	39%	122	28027	82492	89%	54%
63	Aminicenantes	AAA252-K18	SAK	820697	38%	71	19100	46388	90%	35%

64	Aminicenantes	AAA252-L02	SAK	752971	38%	63	17255	59293	89%	14%
67	Aminicenantes	AAA252-L10	SAK	1365566	39%	96	23467	118424	89%	67%
75	Aminicenantes	AAA252-N06	SAK	475639	39%	43	16718	31183	89%	14%
76	Aminicenantes	AAA252-N07	SAK	751034	38%	39	41380	84425	89%	14%
81	Aminicenantes	AAA252-O03	SAK	1065984	40%	90	18744	63680	89%	27%
83	Aminicenantes	AAA252-O09	SAK	1875558	39%	101	33733	168551	89%	62%
84	Aminicenantes	AAA252-O10	SAK	1186657	39%	99	18862	47217	88%	34%
85	Aminicenantes	AAA252-O11	SAK	964336	39%	76	18124	48579	90%	53%
88	Aminicenantes	AAA252-O19	SAK	436443	39%	45	15797	43914	91%	6%
91	Aminicenantes	AAA252-P13	SAK	1038127	39%	86	18142	59923	89%	26%
93	Aminicenantes	AAA252-P19	SAK	1349283	39%	100	26767	87091	89%	43%
119	Aminicenantes	AAA255-A06	SAK	1387242	39%	105	21905	118448	90%	50%
125	Aminicenantes	AAA255-E05	SAK	836730	39%	71	21627	45242	87%	44%
126	Aminicenantes	AAA255-E10	SAK	979257	40%	103	12648	60965	89%	35%
127	Aminicenantes	AAA255-F06	SAK	444384	39%	43	14377	38418	90%	5%
128	Aminicenantes	AAA255-F09	SAK	1005312	39%	74	19850	68838	89%	34%
134	Aminicenantes	AAA255-K14	SAK	275018	39%	31	11511	27127	90%	26%
137	Aminicenantes	AAA255-O15	SAK	431713	38%	34	19852	42712	87%	4%
226	Aminicenantes	0000014-B08	ETL	363946	49%	41	15130	23360	90%	18%
240	Armatimonadetes	0000077-K19	TA	1341701	64%	95	24165	100020	92%	58%
71	Atribacteria	AAA252-M02	SAK	880181	34%	65	20662	39127	83%	37%
124	Atribacteria	AAA255-E04	SAK	152628	33%	20	9971	16596	82%	6%
130	Atribacteria	AAA255-G05	SAK	1653325	35%	131	19172	91556	83%	73%
136	Atribacteria	AAA255-N14	SAK	1424436	35%	117	19645	60434	83%	73%
167	Atribacteria	0000059-I14	TA	498825	35%	27	30727	83442	86%	34%
213	Atribacteria	AB-164-L03	SAK	436823	34%	53	12021	24849	84%	15%
215	Atribacteria	AB-164-P05	SAK	504973	35%	32	27508	54018	82%	14%
216	Atribacteria	AB-164-D21	SAK	975853	35%	78	20253	67295	83%	54%
217	Atribacteria	AB-164-G04	SAK	329112	34%	27	14082	45535	85%	23%
218	Atribacteria	AB-164-M20	SAK	497803	34%	37	21058	44452	84%	30%
219	Atribacteria	AB-164-A22	SAK	939094	35%	83	16684	48554	84%	26%
227	Atribacteria	0000014-F07	ETL	321674	35%	28	15629	51256	83%	9%
231	Atribacteria	0000079-L04	TA	1011276	36%	52	25753	74422	87%	27%
232	Atribacteria	0000079-F20	TA	719583	37%	43	20353	152356	85%	7%
73	BRC1	AAA252-M09	SAK	1174248	58%	113	17182	55280	87%	22%
139	BRC1	AAA257-C11	ETL	2383729	48%	163	30875	99071	88%	20%
171	Caldiserica	0000059-M03	TA	503801	35%	30	24190	41473	92%	29%
25	Calecamantes	AAA471-M6	GBS	1914298	34%	76	50473	158159	88%	82%
282	Calecamantes	0000106-G12	GBS	547191	34%	40	36687	73563	89%	26%
283	Calecamantes	0000106-H18	GBS	590496	34%	54	14690	60983	90%	22%
286	Calecamantes	0000106-I17	GBS	685388	34%	28	25739	63627	88%	26%
287	Calecamantes	0000106-I5	GBS	292729	34%	43	8627	21175	88%	20%
289	Calecamantes	0000106-J16	GBS	1075924	34%	55	23312	93402	89%	65%
290	Calecamantes	0000106-M22	GBS	1075074	34%	48	55946	119269	89%	53%
291	Calecamantes	0000106-N5	GBS	638377	33%	36	35566	94131	89%	26%
292	Calecamantes	0000106-N7	GBS	1477701	34%	52	46686	118668	88%	76%
294	Calecamantes	0000106-P5	GBS	531498	34%	36	26311	67015	88%	19%

37	Cloacimonetes	AAA252-D03	SAK	1017897	37%	85	18221	58508	90%	29%
43	Cloacimonetes	AAA252-E13	SAK	1723610	37%	120	19946	76679	89%	87%
44	Cloacimonetes	AAA252-E20	SAK	1628562	37%	124	22492	71154	89%	70%
45	Cloacimonetes	AAA252-F02	SAK	789155	36%	68	20181	60562	89%	48%
50	Cloacimonetes	AAA252-G07	SAK	499464	38%	60	14411	22012	89%	15%
51	Cloacimonetes	AAA252-G08	SAK	1837784	37%	125	22952	75934	89%	58%
53	Cloacimonetes	AAA252-I11	SAK	1937670	37%	134	27935	87648	90%	85%
56	Cloacimonetes	AAA252-I16	SAK	1965206	37%	114	36762	126073	89%	49%
62	Cloacimonetes	AAA252-K16	SAK	1184434	37%	87	18473	63750	90%	30%
66	Cloacimonetes	AAA252-L05	SAK	753308	37%	71	12591	54275	89%	25%
68	Cloacimonetes	AAA252-L13	SAK	919075	37%	74	18581	52599	90%	27%
72	Cloacimonetes	AAA252-M08	SAK	1697800	37%	108	32307	88617	89%	67%
74	Cloacimonetes	AAA252-N05	SAK	904050	37%	72	16995	49286	89%	30%
77	Cloacimonetes	AAA252-N11	SAK	1150991	37%	94	19947	49882	90%	74%
78	Cloacimonetes	AAA252-N14	SAK	2055867	37%	159	21495	107358	90%	83%
79	Cloacimonetes	AAA252-N17	SAK	2410160	37%	137	35125	101856	90%	99%
80	Cloacimonetes	AAA252-O02	SAK	897531	37%	74	20337	46119	89%	68%
87	Cloacimonetes	AAA252-O17	SAK	1947828	37%	116	33821	77185	89%	85%
89	Cloacimonetes	AAA252-P02	SAK	890714	36%	73	21239	53893	89%	41%
90	Cloacimonetes	AAA252-P10	SAK	2145262	37%	123	36124	78971	89%	94%
92	Cloacimonetes	AAA252-P14	SAK	813448	37%	71	16782	53849	89%	29%
121	Cloacimonetes	AAA255-B16	SAK	1071789	37%	100	16139	62367	89%	58%
123	Cloacimonetes	AAA255-C20	SAK	1363029	37%	120	17519	38030	89%	68%
135	Cloacimonetes	AAA255-L20	SAK	319853	36%	41	10858	32502	88%	18%
148	Cloacimonetes	0000039-A21	TA	399229	37%	29	20465	35154	87%	11%
152	Cloacimonetes	0000039-G13	TA	262919	38%	19	17410	36172	88%	5%
155	Cloacimonetes	0000039-I17	TA	327619	38%	33	14651	33173	88%	20%
159	Cloacimonetes	0000039-M09	TA	516961	37%	44	17216	33791	88%	30%
162	Cloacimonetes	0000039-P09	TA	289645	40%	22	19733	25719	89%	6%
166	Cloacimonetes	0000059-F17	TA	360882	38%	25	24775	57176	88%	14%
170	Cloacimonetes	0000059-L07	TA	712610	37%	50	25413	49049	89%	58%
173	Cloacimonetes	0000059-P03	TA	506680	38%	29	39052	106575	88%	17%
174	Cloacimonetes	0000059-P11	TA	385079	38%	28	21053	43048	88%	21%
273	Cloacimonetes	0000014-D06	ETL	568133	33%	52	14878	25939	92%	7%
274	Cloacimonetes	0000014-D18	ETL	631060	33%	43	25198	44498	92%	14%
275	Cloacimonetes	0000014-H19	ETL	304411	31%	29	16200	26080	86%	24%
276	Cloacimonetes	0000014-K11	ETL	671361	31%	73	12589	30546	88%	16%
199	Diapherotrites	AAA011-K09	HSM	316537	44%	24	17997	50464	91%	10%
206	Diapherotrites	AAA011-E11	HSM	1098042	47%	43	36112	135457	90%	88%
272	Diapherotrites	AAA011-N19	HSM	302064	47%	30	15263	42310	92%	11%
280	EM3	0000106-B10	GBS	148065	33%	7	32978	45922	96%	22%
55	Euryarchaeota	AAA252-I15	SAK	1325681	45%	61	31965	104376	93%	71%
7	Fervidibacteria	AAA471-D06	GBS	206205	51%	18	16027	34339	90%	7%
114	Fervidibacteria	0000001-B9	GBS	811947	51%	74	15793	41092	91%	10%
115	Fervidibacteria	0000001-D5	GBS	586165	50%	46	16772	58009	92%	13%
117	Fervidibacteria	0000001-G10	GBS	1013888	51%	94	17187	33349	92%	46%
175	Fervidibacteria	0000054-H05	GBS	289531	50%	46	8678	20040	90%	6%

2	Crenarchaeota*	AAA471-B05	GBS	831082	28%	34	45166	161140	90%	40%
5	Crenarchaeota*	AAA471-B23	GBS	1110690	29%	58	33318	116584	90%	64%
6	Crenarchaeota*	AAA471-C03	GBS	797737	29%	59	22429	57484	91%	54%
22	Crenarchaeota*	AAA471-L13	GBS	833677	29%	48	30963	81425	91%	68%
23	Crenarchaeota*	AAA471-L14	GBS	940474	29%	63	24388	63737	90%	64%
29	Crenarchaeota*	AAA471-O08	GBS	1149914	29%	24	70010	152928	92%	78%
277	Gracilibacteria	0000069-K10	EPR	759477	23%	65	16910	47111	89%	29%
278	Gracilibacteria	0000069-P22	EPR	341011	22%	30	13152	26435	89%	19%
156	Hydrogenogenetes	0000039-J10	TA	356599	37%	30	15448	25394	91%	13%
172	Hydrogenogenetes	0000059-N10	TA	1200286	37%	72	27261	70263	92%	41%
236	Hydrogenogenetes	0000077-D07	TA	1400936	37%	84	36301	104126	92%	30%
237	Hydrogenogenetes	0000077-O16	TA	1781666	37%	84	36411	97748	92%	74%
32	Latescibacteria	AAA252-B13	SAK	1759976	41%	138	23078	85304	85%	57%
39	Latescibacteria	AAA252-D10	SAK	601658	41%	60	16022	37472	83%	38%
42	Latescibacteria	AAA252-E07	SAK	2290285	42%	164	24533	75516	85%	73%
145	Latescibacteria	AAA257-K07	ETL	1771475	42%	120	23935	97876	86%	23%
146	Marinimicrobia	AAA257-N23	ETL	948616	39%	71	18079	50692	91%	42%
149	Marinimicrobia	0000039-D08	TA	978259	47%	56	27698	61500	94%	63%
150	Marinimicrobia	0000039-E15	TA	493984	47%	30	21162	58668	95%	36%
161	Marinimicrobia	0000039-O11	TA	828046	47%	75	15414	32891	93%	27%
164	Marinimicrobia	0000059-D20	TA	772719	48%	34	39220	88615	93%	47%
165	Marinimicrobia	0000059-E23	TA	618060	47%	41	23061	75054	94%	15%
169	Marinimicrobia	0000059-L03	TA	597650	48%	41	26560	55420	93%	51%
181	Marinimicrobia	AAA011-A05	TA	352761	45%	27	28124	50987	87%	9%
234	Marinimicrobia	0000077-B04	TA	496915	47%	28	23782	66910	93%	12%
259	Marinimicrobia	0000090-C20	TA	566695	48%	46	18905	37004	94%	21%
268	Marinimicrobia	AAA003-E22	TG	888773	37%	53	30580	67111	92%	51%
269	Marinimicrobia	AAA003-L8	TG	1265631	30%	84	32501	138629	96%	62%
777	Marinimicrobia	AAA160-I06	GOM	884929	32%	32	43745	107580	97%	98%
778	Marinimicrobia	AAA298-D23	HOT	979176	31%	10	236086	407043	97%	100%
779	Marinimicrobia	AAA160-C11	GOM	824595	33%	17	55554	190175	97%	86%
780	Marinimicrobia	AAA160-B08	GOM	922550	33%	26	71066	218810	96%	70%
781	Marinimicrobia	AAA076-M08	GOM	392252	33%	18	39976	71263	97%	66%
132	Microgenomates	AAA255-J07	SAK	211711	39%	25	12425	32308	90%	35%
182	Microgenomates	AAA011-E14	HSM	508619	35%	16	60872	89729	93%	66%
193	Microgenomates	AAA011-A19	HSM	594449	35%	30	32651	66188	92%	35%
196	Microgenomates	AAA011-M05	HSM	390930	37%	17	42320	69783	93%	26%
197	Microgenomates	AAA011-B20	HSM	633002	38%	20	40422	98393	96%	62%
204	Microgenomates	AAA011-I21	HSM	363612	50%	30	15880	60210	91%	49%
205	Microgenomates	AAA011-L05	HSM	485374	34%	18	72012	107315	93%	41%
265	Microgenomates	AAA040-P11	EPR	527430	35%	66	11365	51941	91%	52%
266	Microgenomates	AAA011-L6	HSM	946557	39%	46	33427	117241	90%	58%
267	Microgenomates	AAA011-F5	HSM	201385	35%	31	8304	19327	87%	10%
270	Nanoarchaeota	AAA011-D5	HSM	547038	37%	13	135090	273356	90%	40%
192	Nanoarchaeota	AAA011-L22	HSM	348121	32%	27	26462	57369	91%	38%
198	Nanoarchaeota	AAA011-G17	HSM	1021220	32%	38	42021	116957	92%	68%
202	Nanoarchaeota	AAA011-K22	HSM	319835	32%	26	20887	46263	91%	32%

207	Nanoarchaeota	AAA011-J2	HSM	299490	32%	25	18127	51518	92%	18%
208	Nanoarchaeota	AAA011-L15	HSM	204081	33%	17	13811	25074	92%	28%
34	Omnitrophica	AAA252-B19	SAK	518969	42%	54	15140	36581	88%	39%
147	Omnitrophica	AAA257-O07	ETL	767845	46%	60	19198	63036	93%	41%
184	Omnitrophica	AAA011-L16	HSM	468565	43%	35	20271	34579	94%	13%
263	Omnitrophica	AAA011-A17	HSM	2037078	50%	150	20345	75002	94%	80%
100	Parcubacteria	AAA010-E09	HSM	375352	43%	28	22504	35148	87%	38%
138	Parcubacteria	AAA255-P19	SAK	553464	39%	29	51618	70867	93%	70%
178	Parcubacteria	AAA036-E14	EPR	530689	32%	53	14340	33060	90%	53%
179	Parcubacteria	AAA040-L21	EPR	217827	42%	22	14776	30139	86%	38%
185	Parcubacteria	AAA011-A08	HSM	388659	37%	28	28406	58421	86%	53%
189	Parcubacteria	AAA011-A09	HSM	524065	43%	16	64680	152971	90%	65%
191	Parcubacteria	AAA011-J21	HSM	505716	43%	30	26161	43396	93%	42%
209	Parcubacteria	AB-164-E21	SAK	306368	33%	17	34916	53487	91%	38%
262	Parcubacteria	AAA011-N16	HSM	431116	48%	43	17592	42555	90%	36%
255	PER	0000068-E11	EPR	501814	29%	54	13595	43168	91%	36%
233	Synergistetes	0000079-D21	TA	967373	60%	58	27813	58691	93%	72%
221	Thaumarchaeota	AB-179-E04	GBS	291490	46%	20	29390	47905	91%	47%
783	Thaumarchaeota	AAA007-O23	TG	1104502	36%	32	79020	217450	92%	99%
293	Thermotogae	0000106-O11	GBS	343694	39%	33	13496	28587	92%	30%
168	WS1	0000059-K21	TA	509551	65%	54	17441	47351	91%	7%
average of SAGs:				844012	40%	59	26705	69564	90%	40%

Combined assemblies

taxonomic group	OTU	assi	GC	#con	N50	lcon	CDS	comp
Aigarchaeota	1	1180338	61%	128	15605	62478	92%	82%
Aigarchaeota	2	654645	62%	44	24396	64957	92%	54%
Aigarchaeota	3	688171	55%	51	19616	51570	93%	59%
Aigarchaeota	4	617045	60%	44	25175	58479	91%	38%
Aminicenantes	1	2913626	39%	134	42090	174145	88%	100%
Aminicenantes	2	2405693	39%	184	22844	61403	89%	78%
Atribacteria	1	2091422	35%	153	20104	64574	83%	82%
Calescamantes	1	2317902	34%	146	25634	103152	89%	96%
Cloacimonetes	1	3070162	37%	193	26023	160162	90%	100%
Cloacimonetes	2	1842893	37%	154	18073	69396	90%	58%
Cloacimonetes	3	1789375	38%	119	23835	85176	89%	98%
Cloacimonetes	4	611376	37%	60	15751	39081	88%	30%
Cloacimonetes	5	1095584	33%	85	17632	43166	92%	22%
Cloacimonetes	6	839959	31%	90	13155	30011	88%	36%
Fervidibacteria	1	1536863	51%	150	15507	33661	92%	50%
Geoarchaeota	1	1408179	29%	33	87427	208769	91%	89%
Hydrogenedentes	1	2874403	37%	106	48073	166070	92%	98%
Marinimicrobia	1	2357096	47%	101	37642	169513	94%	97%
Marinimicrobia	2	1119009	33%	41	59634	102633	97%	100%
Nanoarchaeota	1	1143895	32%	69	27500	103128	92%	76%
Nanoarchaeota	2	558681	32%	46	22693	53368	92%	49%

average of combined assemblies:	1576967	41%	101	28972	90707	90%	71%
---------------------------------	---------	-----	-----	-------	-------	-----	-----

* SAGs belong to the subgroup Geoarchaeota within the Crenarchaeota.

Supplementary Table 3 | Bacterial and archaeal marker genes used to calculate the TOL.

COGs	Gene name	Gene nb	PhyloSift Marker ID
COG0093	Ribosomal protein L14	1	PMPROK00003
COG0533	Metal-dependent proteases with possible chaperone activity	2	PMPROK00014
COG0092	Ribosomal protein S3	3	PMPROK00015
COG0098	Ribosomal protein S5	4	PMPROK00019
COG0097	Ribosomal protein L6P/L9E	5	PMPROK00020
COG0081	Ribosomal protein L1	6	PMPROK00022
COG0091	Ribosomal protein L22	7	PMPROK00024
COG0185	Ribosomal protein S19	8	PMPROK00025
COG0052	Ribosomal protein S2	9	PMPROK00028
COG0080	Ribosomal protein L11	10	PMPROK00029
COG0541	Signal recognition particle GTPase	11	PMPROK00031
COG0088	Ribosomal protein L4	12	PMPROK00034
COG0186	Ribosomal protein S17	13	PMPROK00041
COG0504	CTP synthase (UTP-ammonia lyase)	14	PMPROK00047
COG0090	Ribosomal protein L2	15	PMPROK00048
COG0184	Ribosomal protein S15P/S13E	16	PMPROK00050
COG0103	Ribosomal protein S9	17	PMPROK00051
COG0256	Ribosomal protein L18	18	PMPROK00052
COG0094	Ribosomal protein L5	19	PMPROK00053
COG0049	Ribosomal protein S7	20	PMPROK00054
TIGR00484	Translation elongation factors (GTPases)	21	PMPROK00060
COG0532	Translation initiation factor 2 (IF-2; GTPase)	22	PMPROK00064
COG0198	Ribosomal protein L24	23	PMPROK00067
COG0100	Ribosomal protein S11	24	PMPROK00068
COG0244	Ribosomal protein L10	25	PMPROK00069
COG0197	Ribosomal protein L16/L10E	26	PMPROK00071
COG0096	Ribosomal protein S8	27	PMPROK00074
COG0087	Ribosomal protein L3	28	PMPROK00075
COG0099	Ribosomal protein S13	29	PMPROK00081
COG0072	Phenylalanyl-tRNA synthetase beta subunit	30	PMPROK00086
COG0016	Phenylalanyl-tRNA synthetase alpha subunit	31	PMPROK00087
COG0200	Ribosomal protein L15	32	PMPROK00092
COG0051	Ribosomal protein S10	33	PMPROK00093
COG0048	Ribosomal protein S12	34	PMPROK00094
COG0149	Triosephosphate isomerase	35	PMPROK00097
COG0343	Queuine/archaeosine tRNA-ribosyltransferase	36	PMPROK00106
COG0127	Xanthosine triphosphate pyrophosphatase	37	PMPROK00123
COG0164	Ribonuclease HII	38	PMPROK00126

Supplementary Table 4 | Phylogenetic inferences and taxa configurations of marker gene trees. (a) Alignments and phylogenetic treeing methods used to calculate marker gene trees. The 15 taxa configurations are the following: (1) all species, comprising 220 archaeal and 2,240 bacterial genomes; (2) reduced species sets as based on the alignment length of equal to or greater than 4000 (gt4k) or 2000 (gt2k) amino acids; (3) reduced species set as based on the number of marker genes present, resulting in 2 groups with equal to or greater than 30 (30g), or 20 marker genes (20g). Abbreviations are: alignment letter (ali); domains (d) of Bacteria (B), Archaea (A), or both (BA); taxa configuration (tax); tree number (tree); tree calculation algorithm (algorithm); number or type of bootstrap replicates (bs), amino acid substitution model (model). **(b)** Results of the phylogenetic inference of taxa groups with bootstrap values for phyla and superphyla. RAxML-light results are not shown since RAxML-light does not calculate bootstrap values. Additional abbreviations are: number of ingroup taxa (#it); total number of taxa in tree (#tit); monophyly count (mc) indicates if group is monophyletic (1) or not monophyletic (0) with calculates monophyly score; Monophyly score = occurrences of monophyly of a group (2 or more taxa) in all bootstrap trees in which this group is present - shown as percentage. Average bootstrap value (average) of all trees considered for monophyly score.

(a)

	ali	d	tax	tree	algorithm	bs	model	tree	algorithm	model	tree	algorithm	bs	model
(1) all species	a	BA	all	1	Fasttree	local	JTT+CAT	16	RaxML-light	GAMMA,LG	31	RaxML	100	GAMMA,JTT
	b	B	all	2	Fasttree	local	JTT+CAT	17	RaxML-light	GAMMA,LG	32	RaxML	100	GAMMA,JTT
	c	A	all	3	Fasttree	local	JTT+CAT	18	RaxML-light	GAMMA,LG	33	RaxML	100	GAMMA,JTT
(2) reduced species (based on alignment length)	d	BA	gt4k	4	Fasttree	local	JTT+CAT	19	RaxML-light	GAMMA,LG	34	RaxML	100	GAMMA,JTT
	e	B	gt4k	5	Fasttree	local	JTT+CAT	20	RaxML-light	GAMMA,LG	35	RaxML	100	GAMMA,JTT
	f	A	gt4k	6	Fasttree	local	JTT+CAT	21	RaxML-light	GAMMA,LG	36	RaxML	100	GAMMA,JTT
	g	BA	gt2k	7	Fasttree	local	JTT+CAT	22	RaxML-light	GAMMA,LG	37	RaxML	100	GAMMA,JTT
	h	B	gt2k	8	Fasttree	local	JTT+CAT	23	RaxML-light	GAMMA,LG	38	RaxML	100	GAMMA,JTT
	i	A	gt2k	9	Fasttree	local	JTT+CAT	24	RaxML-light	GAMMA,LG	39	RaxML	100	GAMMA,JTT
(3) reduced species (based on marker genes present)	j	BA	30g	10	Fasttree	local	JTT+CAT	25	RaxML-light	GAMMA,LG				
	k	B	30g	11	Fasttree	local	JTT+CAT	26	RaxML-light	GAMMA,LG				
	l	A	30g	12	Fasttree	local	JTT+CAT	27	RaxML-light	GAMMA,LG				
	m	BA	20g	13	Fasttree	local	JTT+CAT	28	RaxML-light	GAMMA,LG				
	n	B	20g	14	Fasttree	local	JTT+CAT	29	RaxML-light	GAMMA,LG				
	o	A	20g	15	Fasttree	local	JTT+CAT	30	RaxML-light	GAMMA,LG				

(b)

taxonomic group	bs	#it	tree	d	#tit	tax	algorithm	mc	notes
Acetothermia	100	2	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	2	2	B	2245	all	fasttree	1	
	100	2	31	BA	2460	all	raxml	1	
	98	2	32	B	2245	all	raxml	1	
average	100							100	
Acidobacteria	100	7	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	7	2	B	2245	all	fasttree	1	
	100	7	4	BA	2306	gt4k	fasttree	1	
	100	7	5	B	2095	gt4k	fasttree	1	
	100	7	7	BA	2415	gt2k	fasttree	1	
	100	7	8	B	2177	gt2k	fasttree	1	
	100	7	10	BA	2245	30g	fasttree	1	
	100	7	11	B	2065	30g	fasttree	1	
	100	7	13	BA	2346	20g	fasttree	1	
	100	7	14	B	2146	20g	fasttree	1	
	99	7	31	BA	2460	all	raxml	1	

	97	7	32	B	2245	all	raxml	1	
	100	7	34	BA	2306	gt4k	raxml	1	
	100	7	35	B	2095	gt4k	raxml	1	
	100	7	37	BA	2415	gt2k	raxml	1	
	100	7	38	B	2177	gt2k	raxml	1	
average	100							100	
Actinobacteria	100	214	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	214	2	B	2245	all	fasttree	1	
	100	214	4	BA	2306	gt4k	fasttree	1	
	100	214	5	B	2095	gt4k	fasttree	1	
	100	214	7	BA	2415	gt2k	fasttree	1	
	100	214	8	B	2177	gt2k	fasttree	1	
	100	213	10	BA	2245	30g	fasttree	1	
	100	213	11	B	2065	30g	fasttree	1	
	100	213	13	BA	2346	20g	fasttree	1	
	100	213	14	B	2146	20g	fasttree	1	
	85	214	31	BA	2460	all	raxml	1	
	75	214	32	B	2245	all	raxml	1	
	99	214	34	BA	2306	gt4k	raxml	1	
	85	214	35	B	2095	gt4k	raxml	1	
	91	214	37	BA	2415	gt2k	raxml	1	
	95	214	38	B	2177	gt2k	raxml	1	
average	96							100	
Aenigmarchaeota	100	3	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	3	3	A	220	all	fasttree	1	
	100	2	7	BA	2415	gt2k	fasttree	1	
	100	2	9	A	205	gt2k	fasttree	1	
	99	3	31	BA	2460	all	raxml	1	
	94	3	33	A	220	all	raxml	1	
	100	2	37	BA	2415	gt2k	raxml	1	
	100	2	39	A	205	gt2k	raxml	1	
average	99							100	
Aerophobetes (CD12)	100	4	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	4	2	B	2245	all	fasttree	1	
	100	2	7	BA	2415	gt2k	fasttree	1	
	100	2	8	B	2177	gt2k	fasttree	1	
	100	4	31	BA	2460	all	raxml	1	
	100	4	32	B	2245	all	raxml	1	
	100	2	37	BA	2415	gt2k	raxml	1	
	100	2	38	B	2177	gt2k	raxml	1	
average	100							100	
Aigarchaeota	100	15	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	15	3	A	220	all	fasttree	1	
	100	2	4	BA	2306	gt4k	fasttree	1	
	100	2	6	A	187	gt4k	fasttree	1	
	100	6	7	BA	2415	gt2k	fasttree	1	
	100	6	9	A	205	gt2k	fasttree	1	
	100	4	13	BA	2346	20g	fasttree	1	
	100	4	15	A	200	20g	fasttree	1	
	100	15	31	BA	2460	all	raxml	1	
	100	15	33	A	220	all	raxml	1	
	100	2	34	BA	2306	gt4k	raxml	1	

	100	2	36	A	187	gt4k	raxml	1	
	100	6	37	BA	2415	gt2k	raxml	1	
	100	6	39	A	205	gt2k	raxml	1	
average	100							100	
Aminicenantes (OP8)	100	36	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	36	2	B	2245	all	fasttree	1	
	100	5	4	BA	2306	gt4k	fasttree	1	
	100	5	5	B	2095	gt4k	fasttree	1	
	100	21	7	BA	2415	gt2k	fasttree	1	
	100	21	8	B	2177	gt2k	fasttree	1	
	100	2	10	BA	2245	30g	fasttree	1	
	100	2	11	B	2065	30g	fasttree	1	
	100	9	13	BA	2346	20g	fasttree	1	
	100	9	14	B	2146	20g	fasttree	1	
	100	36	31	BA	2460	all	raxml	1	
	100	36	32	B	2245	all	raxml	1	
	100	5	34	BA	2306	gt4k	raxml	1	
	100	5	35	B	2095	gt4k	raxml	1	
	100	21	37	BA	2415	gt2k	raxml	1	
	100	21	38	B	2177	gt2k	raxml	1	
average	100							100	
Aquificae	100	11	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	11	2	B	2245	all	fasttree	1	
	100	11	4	BA	2306	gt4k	fasttree	1	
	100	11	5	B	2095	gt4k	fasttree	1	
	100	11	7	BA	2415	gt2k	fasttree	1	
	100	11	8	B	2177	gt2k	fasttree	1	
	100	10	10	BA	2245	30g	fasttree	1	
	100	10	11	B	2065	30g	fasttree	1	
	100	11	13	BA	2346	20g	fasttree	1	
	100	11	14	B	2146	20g	fasttree	1	
	100	11	31	BA	2460	all	raxml	1	
	100	11	32	B	2245	all	raxml	1	
	100	11	34	BA	2306	gt4k	raxml	1	
	95	11	35	B	2095	gt4k	raxml	1	
	100	11	37	BA	2415	gt2k	raxml	1	
	99	11	38	B	2177	gt2k	raxml	1	
average	100							100	
Armatimonadetes (OP10)	100	3	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	3	2	B	2245	all	fasttree	1	
	100	3	4	BA	2306	gt4k	fasttree	1	
	100	3	5	B	2095	gt4k	fasttree	1	
	100	3	7	BA	2415	gt2k	fasttree	1	
	100	3	8	B	2177	gt2k	fasttree	1	
	96	3	31	BA	2460	all	raxml	1	
	96	3	32	B	2245	all	raxml	1	
	100	3	34	BA	2306	gt4k	raxml	1	
	100	3	35	B	2095	gt4k	raxml	1	
	100	3	37	BA	2415	gt2k	raxml	1	
	100	3	38	B	2177	gt2k	raxml	1	
average	99							100	
Atribacteria (OP9)	100	14	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	14	2	B	2245	all	fasttree	1	

	100	3	4	BA	2306	gt4k	fasttree	1	
	100	3	5	B	2095	gt4k	fasttree	1	
	100	8	7	BA	2415	gt2k	fasttree	1	
	100	8	8	B	2177	gt2k	fasttree	1	
	100	2	10	BA	2245	30g	fasttree	1	
	100	2	11	B	2065	30g	fasttree	1	
	100	5	13	BA	2346	20g	fasttree	1	
	100	5	14	B	2146	20g	fasttree	1	
	100	14	31	BA	2460	all	raxml	1	
	99	14	32	B	2245	all	raxml	1	
	100	3	34	BA	2306	gt4k	raxml	1	
	100	3	35	B	2095	gt4k	raxml	1	
	99	8	37	BA	2415	gt2k	raxml	1	
	100	8	38	B	2177	gt2k	raxml	1	
average	100							100	
Bacteroidetes	100	112	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	112	2	B	2245	all	fasttree	1	
	100	110	4	BA	2306	gt4k	fasttree	1	
	100	110	5	B	2095	gt4k	fasttree	1	
	100	112	7	BA	2415	gt2k	fasttree	1	
	100	112	8	B	2177	gt2k	fasttree	1	
	100	109	10	BA	2245	30g	fasttree	1	
	100	109	11	B	2065	30g	fasttree	1	
	100	111	13	BA	2346	20g	fasttree	1	
	100	111	14	B	2146	20g	fasttree	1	
	99	112	31	BA	2460	all	raxml	1	
	100	112	32	B	2245	all	raxml	1	
	100	110	34	BA	2306	gt4k	raxml	1	
	100	110	35	B	2095	gt4k	raxml	1	
	100	112	37	BA	2415	gt2k	raxml	1	
	100	112	38	B	2177	gt2k	raxml	1	
average	100							100	
Caldiserica (OP5)	100	2	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	2	2	B	2245	all	fasttree	1	
	93	2	31	BA	2460	all	raxml	1	
	93	2	32	B	2245	all	raxml	1	
average	97							100	
Calescamantes (EM19)	100	10	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	10	2	B	2245	all	fasttree	1	
	100	2	4	BA	2306	gt4k	fasttree	1	
	100	2	5	B	2095	gt4k	fasttree	1	
	100	6	7	BA	2415	gt2k	fasttree	1	
	100	6	8	B	2177	gt2k	fasttree	1	
	100	2	10	BA	2245	30g	fasttree	1	
	100	2	11	B	2065	30g	fasttree	1	
	100	4	13	BA	2346	20g	fasttree	1	
	100	4	14	B	2146	20g	fasttree	1	
	100	10	31	BA	2460	all	raxml	1	
	100	10	32	B	2245	all	raxml	1	
	100	2	34	BA	2306	gt4k	raxml	1	
	100	2	35	B	2095	gt4k	raxml	1	
	100	6	37	BA	2415	gt2k	raxml	1	

	100	6	38	B	2177	gt2k	raxml	1
average	100							100
Chlamydiae	100	17	1	BA	2460	all	fasttree	1
monophyly score = 100	100	17	2	B	2245	all	fasttree	1
	100	17	4	BA	2306	gt4k	fasttree	1
	100	17	5	B	2095	gt4k	fasttree	1
	100	17	7	BA	2415	gt2k	fasttree	1
	100	17	8	B	2177	gt2k	fasttree	1
	100	17	10	BA	2245	30g	fasttree	1
	100	17	11	B	2065	30g	fasttree	1
	100	17	13	BA	2346	20g	fasttree	1
	100	17	14	B	2146	20g	fasttree	1
	100	17	31	BA	2460	all	raxml	1
	100	17	32	B	2245	all	raxml	1
	100	17	34	BA	2306	gt4k	raxml	1
	100	17	35	B	2095	gt4k	raxml	1
	100	17	37	BA	2415	gt2k	raxml	1
	100	17	38	B	2177	gt2k	raxml	1
average	100							100
Chlorobi	100	12	1	BA	2460	all	fasttree	1
monophyly score = 100	100	12	2	B	2245	all	fasttree	1
	100	12	4	BA	2306	gt4k	fasttree	1
	100	12	5	B	2095	gt4k	fasttree	1
	100	12	7	BA	2415	gt2k	fasttree	1
	100	12	8	B	2177	gt2k	fasttree	1
	100	11	10	BA	2245	30g	fasttree	1
	100	11	11	B	2065	30g	fasttree	1
	100	12	13	BA	2346	20g	fasttree	1
	100	12	14	B	2146	20g	fasttree	1
	100	12	31	BA	2460	all	raxml	1
	100	12	32	B	2245	all	raxml	1
	100	12	34	BA	2306	gt4k	raxml	1
	100	12	35	B	2095	gt4k	raxml	1
	100	12	37	BA	2415	gt2k	raxml	1
	100	12	38	B	2177	gt2k	raxml	1
average	100							100
Chloroflexi	100	15	1	BA	2460	all	fasttree	1
monophyly score = 100	100	15	2	B	2245	all	fasttree	1
	100	15	4	BA	2306	gt4k	fasttree	1
	100	15	5	B	2095	gt4k	fasttree	1
	100	15	7	BA	2415	gt2k	fasttree	1
	100	15	8	B	2177	gt2k	fasttree	1
	100	15	10	BA	2245	30g	fasttree	1
	100	15	11	B	2065	30g	fasttree	1
	100	15	13	BA	2346	20g	fasttree	1
	100	15	14	B	2146	20g	fasttree	1
	100	15	31	BA	2460	all	raxml	1
	100	15	32	B	2245	all	raxml	1
	100	15	34	BA	2306	gt4k	raxml	1
	87	15	35	B	2095	gt4k	raxml	1
	100	15	37	BA	2415	gt2k	raxml	1
	99	15	38	B	2177	gt2k	raxml	1
average	99							100

Cloacimonetes (WWE1)	100	38	1	BA	2460	all	fasttree	1
monophyly score = 100	100	38	2	B	2245	all	fasttree	1
	100	14	4	BA	2306	gt4k	fasttree	1
	100	14	5	B	2095	gt4k	fasttree	1
	100	22	7	BA	2415	gt2k	fasttree	1
	100	22	8	B	2177	gt2k	fasttree	1
	100	10	10	BA	2245	30g	fasttree	1
	100	10	11	B	2065	30g	fasttree	1
	100	19	13	BA	2346	20g	fasttree	1
	100	19	14	B	2146	20g	fasttree	1
	100	38	31	BA	2460	all	raxml	1
	99	38	32	B	2245	all	raxml	1
	100	14	34	BA	2306	gt4k	raxml	1
	100	14	35	B	2095	gt4k	raxml	1
	100	22	37	BA	2415	gt2k	raxml	1
	100	22	38	B	2177	gt2k	raxml	1
average	100							100
Cyanobacteria	100	50	1	BA	2460	all	fasttree	1
monophyly score = 100	100	50	2	B	2245	all	fasttree	1
	100	50	4	BA	2306	gt4k	fasttree	1
	100	50	5	B	2095	gt4k	fasttree	1
	100	50	7	BA	2415	gt2k	fasttree	1
	100	50	8	B	2177	gt2k	fasttree	1
	100	50	10	BA	2245	30g	fasttree	1
	100	50	11	B	2065	30g	fasttree	1
	100	50	13	BA	2346	20g	fasttree	1
	100	50	14	B	2146	20g	fasttree	1
	100	50	31	BA	2460	all	raxml	1
	100	50	32	B	2245	all	raxml	1
	100	50	34	BA	2306	gt4k	raxml	1
	100	50	35	B	2095	gt4k	raxml	1
	100	50	37	BA	2415	gt2k	raxml	1
	100	50	38	B	2177	gt2k	raxml	1
average	100							100
Deferribacteres	100	4	1	BA	2460	all	fasttree	1
monophyly score = 100	100	4	2	B	2245	all	fasttree	1
	100	4	4	BA	2306	gt4k	fasttree	1
	100	4	5	B	2095	gt4k	fasttree	1
	100	4	7	BA	2415	gt2k	fasttree	1
	100	4	8	B	2177	gt2k	fasttree	1
	100	4	10	BA	2245	30g	fasttree	1
	100	4	11	B	2065	30g	fasttree	1
	100	4	13	BA	2346	20g	fasttree	1
	100	4	14	B	2146	20g	fasttree	1
	100	4	31	BA	2460	all	raxml	1
	100	4	32	B	2245	all	raxml	1
	100	4	34	BA	2306	gt4k	raxml	1
	100	4	35	B	2095	gt4k	raxml	1
	100	4	37	BA	2415	gt2k	raxml	1
	100	4	38	B	2177	gt2k	raxml	1
average	100							100
Diapherotrites (pMC2A384)	96	2	1	BA	2460	all	fasttree	1
monophyly score = 100	100	3	3	A	220	all	fasttree	1

	70	2	31	BA	2460	all	raxml	1
	74	3	33	A	220	all	raxml	1
average	85							100
Dictyoglomi	100	2	1	BA	2460	all	fasttree	1
monophyly score = 100	100	2	2	B	2245	all	fasttree	1
	100	2	4	BA	2306	gt4k	fasttree	1
	100	2	5	B	2095	gt4k	fasttree	1
	100	2	7	BA	2415	gt2k	fasttree	1
	100	2	8	B	2177	gt2k	fasttree	1
	100	2	10	BA	2245	30g	fasttree	1
	100	2	11	B	2065	30g	fasttree	1
	100	2	13	BA	2346	20g	fasttree	1
	100	2	14	B	2146	20g	fasttree	1
	100	2	31	BA	2460	all	raxml	1
	100	2	32	B	2245	all	raxml	1
	100	2	34	BA	2306	gt4k	raxml	1
	100	2	35	B	2095	gt4k	raxml	1
	100	2	37	BA	2415	gt2k	raxml	1
	100	2	38	B	2177	gt2k	raxml	1
average	100							100
Elusimicrobia	100	2	1	BA	2460	all	fasttree	1
monophyly score = 100	100	2	2	B	2245	all	fasttree	1
	100	2	4	BA	2306	gt4k	fasttree	1
	100	2	5	B	2095	gt4k	fasttree	1
	100	2	7	BA	2415	gt2k	fasttree	1
	100	2	8	B	2177	gt2k	fasttree	1
	100	2	10	BA	2245	30g	fasttree	1
	100	2	11	B	2065	30g	fasttree	1
	100	2	13	BA	2346	20g	fasttree	1
	100	2	14	B	2146	20g	fasttree	1
	100	2	31	BA	2460	all	raxml	1
	100	2	32	B	2245	all	raxml	1
	100	2	34	BA	2306	gt4k	raxml	1
	100	2	35	B	2095	gt4k	raxml	1
	100	2	37	BA	2415	gt2k	raxml	1
	100	2	38	B	2177	gt2k	raxml	1
average	100							100
Epsilonproteobacteria	100	104	1	BA	2460	all	fasttree	1
monophyly score = 100	100	104	2	B	2245	all	fasttree	1
	100	104	4	BA	2306	gt4k	fasttree	1
	100	104	5	B	2095	gt4k	fasttree	1
	100	104	7	BA	2415	gt2k	fasttree	1
	100	104	8	B	2177	gt2k	fasttree	1
	100	103	10	BA	2245	30g	fasttree	1
	100	103	11	B	2065	30g	fasttree	1
	100	104	13	BA	2346	20g	fasttree	1
	100	104	14	B	2146	20g	fasttree	1
	100	104	31	BA	2460	all	raxml	1
	100	104	32	B	2245	all	raxml	1
	100	104	34	BA	2306	gt4k	raxml	1
	100	104	35	B	2095	gt4k	raxml	1
	100	104	37	BA	2415	gt2k	raxml	1
	100	104	38	B	2177	gt2k	raxml	1

	average	100						100	
Euryarchaeota		99	138	1	BA	2460	all	fasttree	1
monophyly score = 100		100	138	3	A	220	all	fasttree	1
		99	137	4	BA	2306	gt4k	fasttree	1
		94	137	6	A	187	gt4k	fasttree	1
		100	138	7	BA	2415	gt2k	fasttree	1
		100	138	9	A	205	gt2k	fasttree	1
		100	137	10	BA	2245	30g	fasttree	1
		93	138	12	A	180	30g	fasttree	1
		100	138	13	BA	2346	20g	fasttree	1
		96	138	15	A	200	20g	fasttree	1
		95	138	31	BA	2460	all	raxml	1
		89	138	33	A	220	all	raxml	1
		97	137	34	BA	2306	gt4k	raxml	1
		96	137	36	A	187	gt4k	raxml	1
		94	138	37	BA	2415	gt2k	raxml	1
		90	138	39	A	205	gt2k	raxml	1
	average	96							100
Fervidibacteria (Oct-Spa1-106)		100	6	1	BA	2460	all	fasttree	1
monophyly score = 100		100	6	2	B	2245	all	fasttree	1
		100	2	7	BA	2415	gt2k	fasttree	1
		100	2	8	B	2177	gt2k	fasttree	1
		85	6	31	BA	2460	all	raxml	1
		81	6	32	B	2245	all	raxml	1
		100	2	37	BA	2415	gt2k	raxml	1
		100	2	38	B	2177	gt2k	raxml	1
	average	96							100
Firmicutes_and_Tenericutes		94	514	1	BA	2460	all	fasttree	0
monophyly score = 63		100	334	2	B	2245	all	fasttree	1
		99	505	4	BA	2306	gt4k	fasttree	0
		99	510	5	B	2095	gt4k	fasttree	0
		100	334	7	BA	2415	gt2k	fasttree	1
		100	334	8	B	2177	gt2k	fasttree	1
		63	505	10	BA	2245	30g	fasttree	0
		61	505	11	B	2065	30g	fasttree	0
		100	509	13	BA	2346	20g	fasttree	0
		100	331	14	B	2146	20g	fasttree	1
		100	216	31	BA	2460	all	raxml	1
		100	216	32	B	2245	all	raxml	1
		100	214	34	BA	2306	gt4k	raxml	1
		100	214	35	B	2095	gt4k	raxml	1
		100	216	37	BA	2415	gt2k	raxml	1
		100	216	38	B	2177	gt2k	raxml	1
	average	95							63
Geoarchaeota		100	7	1	BA	2460	all	fasttree	1
monophyly score = 100		100	7	3	A	220	all	fasttree	1
		100	7	7	BA	2415	gt2k	fasttree	1
		100	7	9	A	205	gt2k	fasttree	1
		100	7	13	BA	2346	20g	fasttree	1
		100	7	15	A	200	20g	fasttree	1
		100	7	31	BA	2460	all	raxml	1
		100	7	33	A	220	all	raxml	1
		100	7	37	BA	2415	gt2k	raxml	1

	100	7	39	A	205	gt2k	raxml	1	
average	100							100	
Gracilibacteria (GN02)	78	2	1	BA	2460	all	fasttree	1	
monophyly score = 100	70	2	2	B	2245	all	fasttree	1	
	100	7	4	BA	2306	gt4k	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	7	7	BA	2415	gt2k	fasttree	1	incl. genome bins by Wrighton et al. 2012
	70	2	8	B	2177	gt2k	fasttree	1	
	100	3	10	BA	2245	30g	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	3	11	B	2065	30g	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	7	13	BA	2346	20g	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	7	14	B	2146	20g	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	2	31	BA	2460	all	raxml	1	
	100	2	32	B	2245	all	raxml	1	
	100	6	34	BA	2306	gt4k	raxml	1	incl. genome bins by Wrighton et al. 2012
	96	7	37	BA	2415	gt2k	raxml	1	incl. genome bins by Wrighton et al. 2012
	100	2	38	B	2177	gt2k	raxml	1	
average	94							100	
Hydrogenedentes (NKB19)	100	4	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	4	2	B	2245	all	fasttree	1	
	100	3	7	BA	2415	gt2k	fasttree	1	
	100	3	8	B	2177	gt2k	fasttree	1	
	100	2	10	BA	2245	30g	fasttree	1	
	100	2	11	B	2065	30g	fasttree	1	
	100	3	13	BA	2346	20g	fasttree	1	
	100	3	14	B	2146	20g	fasttree	1	
	100	4	31	BA	2460	all	raxml	1	
	100	4	32	B	2245	all	raxml	1	
	100	3	37	BA	2415	gt2k	raxml	1	
	100	3	38	B	2177	gt2k	raxml	1	
average	100							100	
Latescibacteria (WS3)	100	4	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	4	2	B	2245	all	fasttree	1	
	100	3	7	BA	2415	gt2k	fasttree	1	
	100	3	8	B	2177	gt2k	fasttree	1	
	100	2	13	BA	2346	20g	fasttree	1	
	100	2	14	B	2146	20g	fasttree	1	
	79	4	31	BA	2460	all	raxml	1	
	70	4	32	B	2245	all	raxml	1	
	100	3	37	BA	2415	gt2k	raxml	1	
	100	3	38	B	2177	gt2k	raxml	1	
average	95							100	
Lentisphaerae	100	2	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	2	2	B	2245	all	fasttree	1	
	100	2	4	BA	2306	gt4k	fasttree	1	
	100	2	5	B	2095	gt4k	fasttree	1	
	100	2	7	BA	2415	gt2k	fasttree	1	
	100	2	8	B	2177	gt2k	fasttree	1	
	100	2	10	BA	2245	30g	fasttree	1	
	100	2	11	B	2065	30g	fasttree	1	
	100	2	13	BA	2346	20g	fasttree	1	
	100	2	14	B	2146	20g	fasttree	1	
	100	2	31	BA	2460	all	raxml	1	
	100	2	32	B	2245	all	raxml	1	

	100	2	34	BA	2306	gt4k	raxml	1
	100	2	35	B	2095	gt4k	raxml	1
	100	2	37	BA	2415	gt2k	raxml	1
	100	2	38	B	2177	gt2k	raxml	1
average	100							100
Marinimicrobia (SAR406, Marine Group A)	100	17	1	BA	2460	all	fasttree	1
monophyly score = 100	100	17	2	B	2245	all	fasttree	1
	100	8	4	BA	2306	gt4k	fasttree	1
	100	8	5	B	2095	gt4k	fasttree	1
	100	13	7	BA	2415	gt2k	fasttree	1
	100	13	8	B	2177	gt2k	fasttree	1
	100	5	10	BA	2245	30g	fasttree	1
	100	5	11	B	2065	30g	fasttree	1
	100	11	13	BA	2346	20g	fasttree	1
	100	11	14	B	2146	20g	fasttree	1
	100	17	31	BA	2460	all	raxml	1
	100	17	32	B	2245	all	raxml	1
	100	8	34	BA	2306	gt4k	raxml	1
	100	8	35	B	2095	gt4k	raxml	1
	100	13	37	BA	2415	gt2k	raxml	1
	100	13	38	B	2177	gt2k	raxml	1
average	100							100
Microgenomates (OP11)	100	11	1	BA	2460	all	fasttree	1
monophyly score = 100	100	11	2	B	2245	all	fasttree	1
	100	8	4	BA	2306	gt4k	fasttree	1
	100	3	5	B	2095	gt4k	fasttree	1
	100	20	7	BA	2415	gt2k	fasttree	1
	100	8	8	B	2177	gt2k	fasttree	1
	100	6	10	BA	2245	30g	fasttree	1
	100	6	11	B	2065	30g	fasttree	1
	100	11	13	BA	2346	20g	fasttree	1
	100	11	14	B	2146	20g	fasttree	1
	100	11	31	BA	2460	all	raxml	1
	100	11	32	B	2245	all	raxml	1
	100	8	34	BA	2306	gt4k	raxml	1
	100	3	35	B	2095	gt4k	raxml	1
	100	20	37	BA	2415	gt2k	raxml	1
	100	8	38	B	2177	gt2k	raxml	1
average	100							100
Nanoarchaeota (Nanoarchaea)	100	6	1	BA	2460	all	fasttree	1
monophyly score = 100	100	7	3	A	220	all	fasttree	1
	100	2	4	BA	2306	gt4k	fasttree	1
	100	2	6	A	187	gt4k	fasttree	1
	100	4	7	BA	2415	gt2k	fasttree	1
	100	4	9	A	205	gt2k	fasttree	1
	93	3	13	BA	2346	20g	fasttree	1
	100	3	15	A	200	20g	fasttree	1
	100	6	31	BA	2460	all	raxml	1
	78	7	33	A	220	all	raxml	1
	78	2	34	BA	2306	gt4k	raxml	1
	78	2	36	A	187	gt4k	raxml	1
	80	4	37	BA	2415	gt2k	raxml	1

	82	4	39	A	205	gt2k	raxml	1	
average	92							100	
Nanohaloarchaeota	100	2	3	A	220	all	fasttree	1	
monophyly score = 100	100	2	4	BA	2306	gt4k	fasttree	1	
	100	2	6	A	187	gt4k	fasttree	1	
	100	2	7	BA	2415	gt2k	fasttree	1	
	100	2	9	A	205	gt2k	fasttree	1	
	100	3	13	BA	2346	20g	fasttree	1	Add. sequence: „Candidatus Haloredivivus“ Ghai et al. 2011
	100	3	15	A	200	20g	fasttree	1	Add. sequence: „Candidatus Haloredivivus“ Ghai et al. 2011
	100	2	33	A	220	all	raxml	1	
	100	2	34	BA	2306	gt4k	raxml	1	
	100	2	36	A	187	gt4k	raxml	1	
	100	2	37	BA	2415	gt2k	raxml	1	
average	100	2	39	A	205	gt2k	raxml	1	
Nitrospirae	100	4	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	4	2	B	2245	all	fasttree	1	
	100	4	4	BA	2306	gt4k	fasttree	1	
	100	4	5	B	2095	gt4k	fasttree	1	
	100	4	7	BA	2415	gt2k	fasttree	1	
	100	4	8	B	2177	gt2k	fasttree	1	
	100	4	10	BA	2245	30g	fasttree	1	
	100	4	11	B	2065	30g	fasttree	1	
	100	4	13	BA	2346	20g	fasttree	1	
	100	4	14	B	2146	20g	fasttree	1	
	92	4	31	BA	2460	all	raxml	1	
	83	4	32	B	2245	all	raxml	1	
	92	4	34	BA	2306	gt4k	raxml	1	
	94	4	35	B	2095	gt4k	raxml	1	
	90	4	37	BA	2415	gt2k	raxml	1	
average	94	4	38	B	2177	gt2k	raxml	1	
Omnitrophica (OP3)	100	4	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	4	2	B	2245	all	fasttree	1	
	100	3	7	BA	2415	gt2k	fasttree	1	
	100	3	8	B	2177	gt2k	fasttree	1	
	100	2	13	BA	2346	20g	fasttree	1	
	100	2	14	B	2146	20g	fasttree	1	
	100	4	31	BA	2460	all	raxml	1	
	99	4	32	B	2245	all	raxml	1	
	97	3	37	BA	2415	gt2k	raxml	1	
average	93	3	38	B	2177	gt2k	raxml	1	
Parcubacteria (OD1)	100	9	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	9	2	B	2245	all	fasttree	1	
	100	15	4	BA	2306	gt4k	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	4	5	B	2095	gt4k	fasttree	1	
	100	22	7	BA	2415	gt2k	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	9	8	B	2177	gt2k	fasttree	1	
	100	7	10	BA	2245	30g	fasttree	1	
	100	7	11	B	2065	30g	fasttree	1	
	100	18	13	BA	2346	20g	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	18	14	B	2146	20g	fasttree	1	incl. genome bins by Wrighton et al. 2012

	100	9	31	BA	2460	all	raxml	1	
	100	9	32	B	2245	all	raxml	1	
	100	15	34	BA	2306	gt4k	raxml	1	incl. genome bins by Wrighton et al. 2012
	100	4	35	B	2095	gt4k	raxml	1	
	100	22	37	BA	2415	gt2k	raxml	1	incl. genome bins by Wrighton et al. 2012
	100	9	38	B	2177	gt2k	raxml	1	
average	100							100	
Parvarchaeota	100	3	3	A	220	all	fasttree	1	
monophyly score = 100	100	2	4	BA	2306	gt4k	fasttree	1	
	100	2	6	A	187	gt4k	fasttree	1	
	100	3	7	BA	2415	gt2k	fasttree	1	
	100	3	9	A	205	gt2k	fasttree	1	
	99	3	13	BA	2346	20g	fasttree	1	
	100	3	15	A	200	20g	fasttree	1	
	82	3	33	A	220	all	raxml	1	
	83	2	34	BA	2306	gt4k	raxml	1	
	84	2	36	A	187	gt4k	raxml	1	
	83	3	37	BA	2415	gt2k	raxml	1	
	86	3	39	A	205	gt2k	raxml	1	
average	93							100	
Planctomycetes	100	9	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	9	2	B	2245	all	fasttree	1	
	100	9	4	BA	2306	gt4k	fasttree	1	
	100	9	5	B	2095	gt4k	fasttree	1	
	100	9	7	BA	2415	gt2k	fasttree	1	
	100	9	8	B	2177	gt2k	fasttree	1	
	100	9	10	BA	2245	30g	fasttree	1	
	100	9	11	B	2065	30g	fasttree	1	
	100	9	13	BA	2346	20g	fasttree	1	
	100	9	14	B	2146	20g	fasttree	1	
	100	9	31	BA	2460	all	raxml	1	
	100	9	32	B	2245	all	raxml	1	
	100	9	34	BA	2306	gt4k	raxml	1	
	100	9	35	B	2095	gt4k	raxml	1	
	100	9	37	BA	2415	gt2k	raxml	1	
	100	9	38	B	2177	gt2k	raxml	1	
average	100							100	
Proteobacteria (A-B-G-D)	86	788	1	BA	2460	all	fasttree	1	
monophyly score = 50%	np	no	2	B	2245	all	fasttree	0	16x Deltaproteobacteria (Desulfovibrionales) cluster outside
	99	772	4	BA	2306	gt4k	fasttree	1	
	no	no	5	B	2095	gt4k	fasttree	0	47x Deltaproteobacteria cluster outside group
	100	783	7	BA	2415	gt2k	fasttree	1	
	100	784	8	B	2177	gt2k	fasttree	1	
	100	762	10	BA	2245	30g	fasttree	1	
	93	759	11	B	2065	30g	fasttree	1	
	83	773	13	BA	2346	20g	fasttree	1	
	87	773	14	B	2146	20g	fasttree	1	
	no	no	31	BA	2460	all	raxml	0	47x Deltaproteobacteria cluster outside group
	no	no	32	B	2245	all	raxml	0	47x Deltaproteobacteria cluster outside group
	no	no	34	BA	2306	gt4k	raxml	0	47x Deltaproteobacteria cluster outside group
	no	no	35	B	2095	gt4k	raxml	0	47x Deltaproteobacteria cluster outside group
	no	no	37	BA	2415	gt2k	raxml	0	47x Deltaproteobacteria cluster outside group

	no	no	38	B	2177	gt2k	raxml	0	47x Deltaproteobacteria cluster outside group
average	94							50	
Spirochaetes	100	61	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	61	2	B	2245	all	fasttree	1	
	100	58	4	BA	2306	gt4k	fasttree	1	
	100	58	5	B	2095	gt4k	fasttree	1	
	100	61	7	BA	2415	gt2k	fasttree	1	
	100	61	8	B	2177	gt2k	fasttree	1	
	100	58	10	BA	2245	30g	fasttree	1	
	100	58	11	B	2065	30g	fasttree	1	
	100	59	13	BA	2346	20g	fasttree	1	
	100	59	14	B	2146	20g	fasttree	1	
	91	61	31	BA	2460	all	raxml	1	
	96	61	32	B	2245	all	raxml	1	
	100	58	34	BA	2306	gt4k	raxml	1	
	100	58	35	B	2095	gt4k	raxml	1	
	99	61	37	BA	2415	gt2k	raxml	1	
	98	61	38	B	2177	gt2k	raxml	1	
average	99							100	
Synergistetes	100	15	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	15	2	B	2245	all	fasttree	1	
	100	15	4	BA	2306	gt4k	fasttree	1	
	100	15	5	B	2095	gt4k	fasttree	1	
	100	15	7	BA	2415	gt2k	fasttree	1	
	100	15	8	B	2177	gt2k	fasttree	1	
	100	14	10	BA	2245	30g	fasttree	1	
	100	14	11	B	2065	30g	fasttree	1	
	100	15	13	BA	2346	20g	fasttree	1	
	100	15	14	B	2146	20g	fasttree	1	
	100	15	31	BA	2460	all	raxml	1	
	100	15	32	B	2245	all	raxml	1	
	100	15	34	BA	2306	gt4k	raxml	1	
	100	15	35	B	2095	gt4k	raxml	1	
	100	15	37	BA	2415	gt2k	raxml	1	
	100	15	38	B	2177	gt2k	raxml	1	
average	100							100	
Thaumarchaeota	100	7	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	7	3	A	220	all	fasttree	1	
	100	5	4	BA	2306	gt4k	fasttree	1	
	100	5	6	A	187	gt4k	fasttree	1	
	100	7	7	BA	2415	gt2k	fasttree	1	
	100	7	9	A	205	gt2k	fasttree	1	
	100	5	10	BA	2245	30g	fasttree	1	
	100	5	12	A	180	30g	fasttree	1	
	100	6	13	BA	2346	20g	fasttree	1	
	100	6	15	A	200	20g	fasttree	1	
	100	7	31	BA	2460	all	raxml	1	
	100	7	33	A	220	all	raxml	1	
	100	5	34	BA	2306	gt4k	raxml	1	
	100	5	36	A	187	gt4k	raxml	1	
	100	7	37	BA	2415	gt2k	raxml	1	
	100	7	39	A	205	gt2k	raxml	1	
average	100							100	

Thermi (Deinococcus-Thermus)	100	19	1	BA	2460	all	fasttree	1
monophyly score = 100	100	19	2	B	2245	all	fasttree	1
	100	19	4	BA	2306	gt4k	fasttree	1
	100	19	5	B	2095	gt4k	fasttree	1
	100	19	7	BA	2415	gt2k	fasttree	1
	100	19	8	B	2177	gt2k	fasttree	1
	100	19	10	BA	2245	30g	fasttree	1
	100	19	11	B	2065	30g	fasttree	1
	100	19	13	BA	2346	20g	fasttree	1
	100	19	14	B	2146	20g	fasttree	1
	100	19	31	BA	2460	all	raxml	1
	100	19	32	B	2245	all	raxml	1
	100	19	34	BA	2306	gt4k	raxml	1
	100	19	35	B	2095	gt4k	raxml	1
	100	19	37	BA	2415	gt2k	raxml	1
	100	19	38	B	2177	gt2k	raxml	1
average	100							100
Thermodesulfobacteria	100	3	1	BA	2460	all	fasttree	1
monophyly score = 100	100	3	2	B	2245	all	fasttree	1
	100	3	4	BA	2306	gt4k	fasttree	1
	100	3	5	B	2095	gt4k	fasttree	1
	100	3	7	BA	2415	gt2k	fasttree	1
	100	3	8	B	2177	gt2k	fasttree	1
	100	3	10	BA	2245	30g	fasttree	1
	100	3	11	B	2065	30g	fasttree	1
	100	3	13	BA	2346	20g	fasttree	1
	100	3	14	B	2146	20g	fasttree	1
	100	3	31	BA	2460	all	raxml	1
	100	3	32	B	2245	all	raxml	1
	100	3	34	BA	2306	gt4k	raxml	1
	100	3	35	B	2095	gt4k	raxml	1
	100	3	37	BA	2415	gt2k	raxml	1
	100	3	38	B	2177	gt2k	raxml	1
average	100							100
Thermoprotei	100	21	1	BA	2460	all	fasttree	1
monophyly score = 100	100	21	3	A	220	all	fasttree	1
	100	21	4	BA	2306	gt4k	fasttree	1
	100	21	6	A	187	gt4k	fasttree	1
	100	21	7	BA	2415	gt2k	fasttree	1
	100	21	9	A	205	gt2k	fasttree	1
	100	21	10	BA	2245	30g	fasttree	1
	100	21	12	A	180	30g	fasttree	1
	100	21	13	BA	2346	20g	fasttree	1
	100	21	15	A	200	20g	fasttree	1
	100	21	31	BA	2460	all	raxml	1
	100	21	33	A	220	all	raxml	1
	100	21	34	BA	2306	gt4k	raxml	1
	100	21	36	A	187	gt4k	raxml	1
	100	21	37	BA	2415	gt2k	raxml	1
	100	21	39	A	205	gt2k	raxml	1
average	100							100
Thermotogae	100	15	1	BA	2460	all	fasttree	1
monophyly score = 100	100	15	2	B	2245	all	fasttree	1

	100	14	4	BA	2306	gt4k	fasttree	1	
	100	14	5	B	2095	gt4k	fasttree	1	
	100	15	7	BA	2415	gt2k	fasttree	1	
	100	15	8	B	2177	gt2k	fasttree	1	
	100	14	10	BA	2245	30g	fasttree	1	
	100	14	11	B	2065	30g	fasttree	1	
	100	15	13	BA	2346	20g	fasttree	1	
	100	15	14	B	2146	20g	fasttree	1	
	100	15	31	BA	2460	all	raxml	1	
	100	15	32	B	2245	all	raxml	1	
	100	14	34	BA	2306	gt4k	raxml	1	
	100	14	35	B	2095	gt4k	raxml	1	
	100	15	37	BA	2415	gt2k	raxml	1	
	100	15	38	B	2177	gt2k	raxml	1	
average	100							100	
Verrucomicrobia	100	9	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	9	2	B	2245	all	fasttree	1	
	100	9	4	BA	2306	gt4k	fasttree	1	
	100	9	5	B	2095	gt4k	fasttree	1	
	100	9	7	BA	2415	gt2k	fasttree	1	
	100	9	8	B	2177	gt2k	fasttree	1	
	100	9	10	BA	2245	30g	fasttree	1	
	100	9	11	B	2065	30g	fasttree	1	
	100	9	13	BA	2346	20g	fasttree	1	
	100	9	14	B	2146	20g	fasttree	1	
	100	9	31	BA	2460	all	raxml	1	
	100	9	32	B	2245	all	raxml	1	
	100	9	34	BA	2306	gt4k	raxml	1	
	100	9	35	B	2095	gt4k	raxml	1	
	100	9	37	BA	2415	gt2k	raxml	1	
	100	9	38	B	2177	gt2k	raxml	1	
average	100							100	
SUPERPHYLA									
TACK	100	63	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	64	3	A	220	all	fasttree	1	
	100	43	4	BA	2306	gt4k	fasttree	1	
	100	43	6	A	187	gt4k	fasttree	1	
	100	55	7	BA	2415	gt2k	fasttree	1	
	100	55	9	A	205	gt2k	fasttree	1	
	100	41	10	BA	2245	30g	fasttree	1	
	100	41	12	A	180	30g	fasttree	1	
	100	52	13	BA	2346	20g	fasttree	1	
	100	52	15	A	200	20g	fasttree	1	
	100	63	31	BA	2460	all	raxml	1	
	100	64	33	A	220	all	raxml	1	
	100	43	34	BA	2306	gt4k	raxml	1	
	100	43	36	A	187	gt4k	raxml	1	
	100	55	37	BA	2415	gt2k	raxml	1	
	100	55	39	A	205	gt2k	raxml	1	
average	100							100	
FCB	100	188	1	BA	2460	all	fasttree	1	
monophyly score = 100	100	188	2	B	2245	all	fasttree	1	
	100	149	4	BA	2306	gt4k	fasttree	1	

	100	149	5	B	2095	gt4k	fasttree	1	
	100	166	7	BA	2415	gt2k	fasttree	1	
	100	166	8	B	2177	gt2k	fasttree	1	
	100	139	10	BA	2245	30g	fasttree	1	
	100	139	11	B	2065	30g	fasttree	1	
	100	159	13	BA	2346	20g	fasttree	1	
	100	159	14	B	2146	20g	fasttree	1	
	78	188	31	BA	2460	all	raxml	1	
	67	188	32	B	2245	all	raxml	1	
	99	149	34	BA	2306	gt4k	raxml	1	
	99	149	35	B	2095	gt4k	raxml	1	
	100	166	37	BA	2415	gt2k	raxml	1	
	98	166	38	B	2177	gt2k	raxml	1	
average	96							100	
DPANN	no	no	1	BA	2460	all	fasttree	0	Diapherotrites outside group (pulled towards Bacteria)
monophyly score = 56	100	18	3	A	220	all	fasttree	1	
	no	no	4	BA	2306	gt4k	fasttree	0	Nanohaloarchaea outside group (pulled towards Bacteria)
	100	7	6	A	187	gt4k	fasttree	1	
	no	no	7	BA	2415	gt2k	fasttree	0	Nanohaloarchaea outside group (pulled towards Bacteria)
	100	12	9	A	205	gt2k	fasttree	1	
	100	2	10	BA	2245	30g	fasttree	1	
	100	2	12	A	180	30g	fasttree	1	
	no	no	13	BA	2346	20g	fasttree	0	group splits into several subgroups
	100	10	15	A	200	20g	fasttree	1	
	no	no	31	BA	2460	all	raxml	0	group splits into several subgroups
	100	18	33	A	220	all	raxml	1	
	no	no	34	BA	2306	gt4k	raxml	0	group splits into several subgroups
	100	7	36	A	187	gt4k	raxml	1	
	no	no	37	BA	2415	gt2k	raxml	0	group splits into several subgroups
	99	12	39	A	205	gt2k	raxml	1	
average	100							56	
Patescibacteria	100	24	1	BA	2460	all	fasttree	1	
monophyly score = 100	97	24	2	B	2245	all	fasttree	1	
	100	33	4	BA	2306	gt4k	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	8	5	B	2095	gt4k	fasttree	1	
	100	55	7	BA	2415	gt2k	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	21	8	B	2177	gt2k	fasttree	1	
	100	21	10	BA	2245	30g	fasttree	1	
	100	21	11	B	2065	30g	fasttree	1	
	100	46	13	BA	2346	20g	fasttree	1	incl. genome bins by Wrighton et al. 2012
	100	46	14	B	2146	20g	fasttree	1	incl. genome bins by Wrighton et al. 2012
	82	24	31	BA	2460	all	raxml	1	
	90	24	32	B	2245	all	raxml	1	
	79	33	34	BA	2306	gt4k	raxml	1	incl. genome bins by Wrighton et al. 2012
	100	8	35	B	2095	gt4k	raxml	1	
	95	55	37	BA	2415	gt2k	raxml	1	incl. genome bins by Wrighton et al. 2012
	91	21	38	B	2177	gt2k	raxml	1	
average	96							100	
PVC	no	no	1	BA	2460	all	fasttree	0	group splits into several subgroups
monophyly score = 94	100	41	2	B	2245	all	fasttree	1	
	100	38	4	BA	2306	gt4k	fasttree	1	

	100	37	5	B	2095	gt4k	fasttree	1	
	100	40	7	BA	2415	gt2k	fasttree	1	
	96	40	8	B	2177	gt2k	fasttree	1	
	100	37	10	BA	2245	30g	fasttree	1	
	100	37	11	B	2065	30g	fasttree	1	
	100	39	13	BA	2346	20g	fasttree	1	
	100	39	14	B	2146	20g	fasttree	1	
	82	41	31	BA	2460	all	raxml	1	
	94	41	32	B	2245	all	raxml	1	
	99	38	34	BA	2306	gt4k	raxml	1	
	99	38	35	B	2095	gt4k	raxml	1	
	89	40	37	BA	2415	gt2k	raxml	1	
	89	40	38	B	2177	gt2k	raxml	1	
average	97							94	
Terrabacteria	76	887	1	BA	2460	all	fasttree	1	
monophyly score = 63	100	887	2	B	2245	all	fasttree	1	
	100	814	4	BA	2306	gt4k	fasttree	0	OP9 clusters within Terrabacteria
	100	904	5	B	2095	gt4k	fasttree	0	OP9 clusters within Terrabacteria
	100	887	7	BA	2415	gt2k	fasttree	1	
	65	887	8	B	2177	gt2k	fasttree	1	
	96	890	10	BA	2245	30g	fasttree	0	OP9 clusters within Terrabacteria
	97	890	11	B	2065	30g	fasttree	0	OP9 clusters within Terrabacteria
	100	807	13	BA	2346	20g	fasttree	0	Tenericutes and Erysipelotrichales (Firmicutes) cluster outside
	100	879	14	B	2146	20g	fasttree	1	
	83	870	31	BA	2460	all	raxml	1	
	22	869	32	B	2245	all	raxml	1	
	55	865	34	BA	2306	gt4k	raxml	1	
	99	38	35	B	2095	gt4k	raxml	0	Patescibacteria cluster within group
	53	869	37	BA	2415	gt2k	raxml	1	
	53	869	38	B	2177	gt2k	raxml	1	
average	81							63	

Wrighton, K. C. et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* 337, 1661–1665 (2012).

Supplementary Table 5 | Proposed names for candidate phyla and associated superphyla. For each candidatus type species the sampling site (see also Fig. 1), Microbial Dark Matter project ID, name, assembly size (Mbp), estimated genome completeness (%), and IMG taxon ID are provided in parenthesis. See supplementary Table 2 for detailed genome statistics.

Super-phylum	Candidate phylum	Proposed name and etymology	Candidatus type species (existing <i>Candidatus</i> species are referenced)	#SAGs (#total)
PVC	OP3	“Omnitrophica” Om.ni.tro’phi.ca. A higher taxonomic unit comprising the genus <i>Omnitrophus</i>	<i>Candidatus Omnitrophus fodinae</i> (HSM, MDM ID 263, SCGC AAA011-A17, 2.0 Mbp, 79.9%, 2264867162) Om.ni.tro’phus. L. adj. omnis, all; N.L. n. trophus (from Gr. trepho, to eat), eating; N.L. masc. n. Omnitrophus, eating all fo.di’nae. L. gen. n. fodinae, of (from) a mine, coal mine	4 (4)
	SAR406 (Marine Group A)	“Marinimicrobia” Ma.ri.ni.mi.cro’bi.a. A higher taxonomic unit comprising the genus <i>Marinimicrobium</i>	<i>Candidatus Marinimicrobium atlanticum</i> (GOM, combined assembly Marinimicrobia -2, 1.1 Mbp, 100%, 2518285581) Ma.ri.ni.mi.cro’bi.um. L. adj. marinus, marine; N.L. n. microbium, a small living entity; N.L. neutr. n. Marinimicrobium, a marine microbe. at.lan’ti.cum. N.L. adj. atlanticum referring to the Atlantic ocean from where the sample was taken	17 (17)
FCB	WS3	“Latescibacteria” La.tes.ci.bac.te’ri.a. A higher taxonomic unit comprising the genus <i>Latescibacter</i>	<i>Candidatus Latescibacter anaerobius</i> (SAK, MDM ID 42, SCGC AAA252-E07, 2.3Mbp, 72.7%, 2264867254) La.tes.ci.bac’ter. L. v. latesco, to hide; -i- connecting vowel; N. L. n. bacter, a small rod; N.L. masc. n. Latescibacter a hiding small rod. an.a.er.o’bi.us. Gr. pref. an- negating; Gr. n. aer, air; Gr. n. bios life; N.L. adj. anaerobius living without air (i. e. oxygen)	4 (4)
	WWE1	“Cloacimonetes” Clo.a.ci.mo.ne’tes. A higher taxonomic unit comprising the genus <i>Cloacimonas</i> .	<i>Candidatus Cloacimonas acidaminovorans</i>¹ † Clo.a.ci.mo’nas. L. n. cloaca, sewer; -i- connecting vowel; N.L. n. monas a unit, N.L. fem. n. Cloacimonas a unit from a sewer a.cid.a.mi.no.vo’rans. L. n. acidum an acid; N.L. n. aminum an amine; L. part. adj. vorans eating. N.L. adj. acidaminovorans, amino acid eating.	37 (38)
	OP8	“Aminicenantes” A.mi.ni.ce.nan’tes. A higher taxonomic unit comprising the genus <i>Aminicenans</i> .	<i>Candidatus Aminicenans sakinaicola</i> (SAK, combined assembly Aminicenantes -1, 3.0 Mbp, 100%, 2518285573) A.mi.ni.ce’nans. N.L. n. aminum, amine; L. v. cenare, to eat; N.L. masc. n. Aminicenans a (bacterium) degrading amino acids -sa.ki.na.wi’co.la. N. L. n. sakinaicola referring to the sampling site, Lake Sakinaicola; -i- connecting vowel; L. suffix -cola, a dweller; N.L. n. sakinaicola dwelling in Lake Sakinaicola	36 (36)
Patescibacteria† -patesco (lat.), bare	OP11	“Microgenomates” A higher taxonomic unit comprising the genus <i>Microgenomatus</i>	<i>Candidatus Microgenomatus auricola</i> (HSM, MDM ID 182, SCGC AAA011-E14, 0.5 Mbp, 65.5%, 2264867148) Mi.cro.ge.no.ma’tus. Gr. Adj. mikros, small; N.L. n. genomum, genome; N.L. n. Microgenomatus; an organism with a small genome size (~1 Mbp) au.ri’co.la. L. n. aurum, gold; L. suff. -cola, a dweller; N.L. n. auricola, a dweller on gold	10 (22)
	OD1	“Parcubacteria” -parcus (lat.), thrifty A higher taxonomic unit comprising the genus <i>Paceibacter</i>	<i>Candidatus Paceibacter normanii</i> (SAK, MDM ID 138, SCGC AAA255-P19, 0.6 Mbp, 70.3%, 2264867004) Pa.ce.i.bac’ter. Pace referring to Norman Pace. N.L. n. bacter (from Gr. bakterion) a small rod. N.L. masc. n. Paceibacter Pace’s bacterium norman’i.i. N.L. gen. N. of Norman, referring to Norman Pace (Norman Richard Pace, Jr. is an American biochemist, Distinguished Professor of Molecular, Cellular and Developmental Biology at the University of Colorado, and principal investigator at the Pace lab)	9 (23)
	GN02 (BD1-5)	“Gracilibacteria” -gracilis (lat.), slim, slender, slight, meager, simple A higher taxonomic unit comprising the genus <i>Altamarinus</i>	<i>Candidatus Altamarinus pacificus</i> (EPR, MDM ID 255, JGI 0000068-E11, 0.5 Mbp, 36%, 2264867094) Al.ti.ma.ri’nus. L. adj. altus, high, deep; L. adj. marinus, marine; N. L. masc. n. Altamarinus a (bacterium) from the Deep Sea pa.ci’fi.cus. N.L. masc. adj. referring to the Pacific Ocean.	2 (7)

	OP9	“Atribacteria” A.tri.bac.te’ri.a. N.L. n. A higher taxonomic unit comprising the genus <i>Caldatribacterium</i>	<i>Candidatus Caldatribacterium californiense</i>² Cald.at.ri.bac.te’ri.um. L. adj. caldus, hot; L. adj. ater, black, dark; N.L. n. bacterium, a small rod; N.L. neutr. n. Caldatribacterium a small rod from hot dark matter ca.li.for.ni.en’s.e. N.L. adj. of/from the state California	14 (16)
	EM19	“Calescamantes” Ca.lesc.a.man’tes. L. v. calesco, to become warm, grow hot; L. v. amo, to love, N.L. n. Calescamantes heat lovers A higher taxonomic unit comprising the genus <i>Calescibacterium</i>	<i>Candidatus Calescibacterium nevadense</i> (GBS, combined assembly Calescamantes, 2.2 Mbp, 92%, 2518285568) Cal.es.ci.bac.te’ri.um. L. v. calesco, to become warm, grow hot; N.L. n. bacterium a small rod. N.L. n. Calescibacterium a bacterium from an extremely hot environment. ne.va.den’s.e. N.L. n. nevadense of/from the state of Nevada	10 (1)
	CD12 (BHI80-139)	“Aerophobetes” A higher taxonomic unit comprising the genus <i>Aerophobus</i>	<i>Candidatus Aerophobus profundus</i> (SAK, MDM ID 70, SCGC AAA252-L23, 2.3 Mbp, 67.1%, 2264867061) A.er.o.pho’bus. Gr. n. aer, air; Gr. adj. phobos, fear. N. L. n. Aerophobus, fearing of air (i. e. oxygen). pro.fun’dus. L. adj. profundus, deep	4 (4)
	NKB19	“Hydrogenedentes” A higher taxonomic unit comprising the genus <i>Hydrogenedens</i>	<i>Candidatus Hydrogenedens terephthalicus</i> (TA, combined assembly Hydrogenedentes-1, 2.9 Mbp, 96.7%, 2518285572) Hy.dro.ge.ne’dens. N.L.n. hydrogenium, hydrogen; L. V. edo, eat; N.L. n. Hydrogenedens a hydrogen consumer te.re.phtha’li.cus. N.L. n. (acidum) terephthalicum, terephthalic acid. N.L. adj. terephthalicus, referring to the environment of isolation, a terephthalate reactor.	4 (4)
	OP1	“Acetothermia” A.ce.to.ther’mi.a. A higher taxonomic unit comprising the genus <i>Acetothermum</i>	<i>Candidatus Acetothermum autotrophicum</i>³ ‘Acetothermus’ (A.ce.to.ther’mus. L.n. acetum vinegar; Gr. Adj. thermos hot; M.L. masc. n. Acetothermus indicates a vinegar organism living in hot places.) and ‘autotrophicum’ (au.to.tro’phi.cum. Gr. pron.autos self; Gr. adj. trophikos nursing, tending or feeding; N.L. neut. adj. autotrophicum self- nursing or self-feeding).	1 (2)
	Oct-Spa1- 106	“Fervidibacteria” Fer.vi.di.bac.te’ri.a. A higher taxonomic unit comprising the genus <i>Fervidibacter</i>	<i>Candidatus Fervidibacter sacchari</i> (GBS, MDM ID 117, JGI_0000001-G10, 1.0 Mbp, 45.6%, 2264867076) Fer.vi.do.bac’ter. L. adj. fervidus, hot, steaming; -i- connecting vowel; N.L. n. bacter, a rod; N.L. n. Fervidibacter a hot rod. sac.cha’ri. N.L. n. saccharum, sugar; N.L. gen. n. sacchari, of sugar	5 (5)
DPANN	pMC2A 384	“Diapherotrites” (Dia.phe.ro.tri’tes. Gr. v. diaphero, to differ; Gr. adj. tritos the third; N. L. n. Diapherotrites the 3rd HSM group discovered) A higher taxonomic unit comprising the genus <i>Iainarchaeum</i>	<i>Candidatus Iainarchaeum andersonii</i> (HSM, MDM ID 206, SCGC_AAA011-E11, 1.1 Mbp, 88.5%, 2264867067) (Ia.in.ar.chae’um. Iain, first name of Iain Anderson (in memoriam of Iain Anderson, a Ph.D. in biochemistry from the University of Wisconsin- Madison and a key member of IMG team). N.L. n. archaeum a taxonomic unit. N.L. neutr. n. Iainarchaeum, Iain’s archaeum.) (an.der.so’ni.i. N.L. gen. n. andersonii, of Anderson	3 (3)
	ARMAN group	“Parvarchaeota” Parv.ar.chae’a. A higher taxonomic unit comprising the genera <i>Parvarchaeum</i> and <i>Micrarchaeum</i>	<i>Candidatus Parvarchaeum acidiphilum (ARMAN-4)</i>⁴ Parv.ar.chae’um. L. adj. parvus small; N.L. n. archaeum, a taxonomic unit; N.L. n. Parvarchaeum the small archaeum a.ci.di’phi.lum. L. n. acidum acid, N.L. adj. philus (from Gr. philos) friend, loving; N.L. neutr. adj. acidiphilum acid-loving	0 (3)
	DSEG	“Aenigmarchaeota” Ae.nig.mar.chae’a. A higher taxonomic unit comprising the genus <i>Aenigmarchaeum</i>	<i>Candidatus Aenigmarchaeum subterraneum</i> (HSM, MDM ID 271, AAA011-O16, 0.42 Mbp, 47.3%, 2264867065) Ae.nig.mar.chae’um. L. n. aenigma, riddle; N.L. n. archaeum, a taxonomic unit; N.L. neutr. n. Aenigmarchaeum the enigmatic archaeum. Sub.terr.a’ne.um. L. neutr. adj. subterraneum underground, subterranean	3 (3)
		“Nanohaloarchaeota” (Na.no.ha.lo.ar.chae.o’ta. Gr. n. nanos, dwarf; Gr. n. hal, - los, salt; Gr. adj. archaios, old;	<i>Candidatus Nanosalina sp. J07AB43</i>⁵ Na.no.sa.li’na.. Gr. n. nanos, dwarf; L. n. salina, a salt-work, a place where salt is made N. L. n. Nanosalina a small archaeum from a place where salt is made	0 (2)

		N.L. suffix -ota ending to design a phylum; N.L. n. Nanohaloarchaeota, small salt-loving Archaeota. A higher taxonomic unit comprising the genus <i>Nanosalina</i>	
--	--	--	--

¹ Pelletier, E. *et al.* ‘Candidatus Cloacamonas Acidaminovorans’: Genome Sequence Reconstruction Provides a First Glimpse of a New Bacterial Division. *J. Bacteriol.* **190**, 2572–2579 (2008).

² Dodsworth, J. *et al.* Single-cell and metagenomic analyses indicate a fermentative, saccharolytic lifestyle for members of the OP9 lineage. *Nature Communications*. **Passed pre-review**.

³ Takami, H. *et al.* A Deeply Branching Thermophilic Bacterium with an Ancient Acetyl-CoA Pathway Dominates a Subsurface Ecosystem. *PLoS ONE* **7**, e30559 (2012).

⁴ Baker, B. J. *et al.* Enigmatic, ultrasmall, uncultivated Archaea. *PNAS* **107**, 8806–8811 (2010).

⁵ Narasingarao, P. *et al.* De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J* **6**, 81–93 (2012).

† We update the naming of *Candidatus* Cloacimonas acidaminovorans to honor the Latin origin which requires -i- as a connecting vowel, according to the naming convention for prokaryotes⁶⁰.

‡ See also Supplementary Fig. 8 for evidence of a possible association of the Patescibacteria with the Terrabacteria.

Supplementary Table 6 | IMG annotations of genes involved in energy metabolism and autotrophy.

	category	MDM ID	SAG name	taxonomic group	gene Object ID	locustag	gene product name (updated)	details	(bp)	(aa)
Autotroph	RuBisCO	288	JGI_0000106-J15	Aigarchaeota	2265093640	J0000106J15DRAFT_00921	Ribulose 1,5-bisphosphate carboxylase, large subunit (EC:4.1.1.39)	RuBisCO type III like	1290	429
	RuBisCO	199	SCGC_AAA011-K09	Diapherotrites	2264948351	A11K09DRAFT_00043	ribulose bisphosphate carboxylase, type III (EC:4.1.1.39)	RuBisCO type III	1302	433
	RuBisCO	55	SCGC_AAA252-I15	Euryarchaeota	2265081141	A252I15DRAFT_00086	ribulose bisphosphate carboxylase, type III (EC:4.1.1.39)	RuBisCO type III	1272	423
	RuBisCO	42	SCGC_AAA252-E07	Latescibacteria	2265113049	A252E7DRAFT_00526	Ribulose 1,5-bisphosphate carboxylase, large subunit (EC:4.1.1.39)	RuBisCO type III	1416	471
	RuBisCO	270	SCGC_AAA011-D5	Nanoarchaeota	2264947191	A11D05DRAFT_00533	ribulose bisphosphate carboxylase, type III (EC:4.1.1.39)	RuBisCO type III	1302	433
	RuBisCO	198	SCGC_AAA011-G17	Nanoarchaeota	2264951958	A11G17DRAFT_01137	Ribulose 1,5-bisphosphate carboxylase, large subunit (EC:4.1.1.39)	RuBisCO type III	1224	407
	RuBisCO	209	SCGC_AB-164-E21	Parcubacteria	2264967912	AB164E21DRAFT_00184	ribulose bisphosphate carboxylase, type III (EC:4.1.1.39)	RuBisCO type III	1266	421
	Reductive acetyl-CoA pathway	222	JGI_0000014-C22	Aerophobetes	2264943080	J0000014C22DRAFT_00188	CO dehydrogenase/CO-methylating acetyl-CoA synthase complex, beta subunit (EC:1.2.7.4, EC:1.2.99.2, EC:2.3.1.169)	CO dehydrogenase/acetyl-CoA synthase	2217	738
	Reductive acetyl-CoA pathway	129	SCGC_AAA255-F10	Aerophobetes	2264941872	A255F10DRAFT_00340	CO dehydrogenase/CO-methylating acetyl-CoA synthase complex, beta subunit (EC:2.3.1.169, EC:1.2.7.4, EC:1.2.99.2)	CO dehydrogenase/acetyl-CoA synthase	2217	738
	Reductive acetyl-CoA pathway	129	SCGC_AAA255-F10	Aerophobetes	2264941871	A255F10DRAFT_00339	CO dehydrogenase/acetyl-CoA synthase gamma subunit (corrinoid Fe-S protein) (EC:2.1.1.-)	CO dehydrogenase/acetyl-CoA synthase	1353	450
	Reductive acetyl-CoA pathway	184	SCGC_AAA011-L16	Omnitrophica	2265028498	A11L16DRAFT_00121	CO dehydrogenase/CO-methylating acetyl-CoA synthase complex, beta subunit (EC:1.2.99.2, EC:2.3.1.169, EC:1.2.7.4)	CO dehydrogenase/acetyl-CoA synthase	2214	737
	Reductive acetyl-CoA pathway	276	JGI_0000014-K11	Cloacimonetes	2265119478	J0000014K11DRAFT_00164	CO dehydrogenase/CO-methylating acetyl-CoA synthase complex, beta subunit (EC:2.3.1.169, EC:1.2.99.2, EC:1.2.7.4)	CO dehydrogenase/acetyl-CoA synthase	2160	719
	Reductive acetyl-CoA pathway	55	SCGC_AAA252-I15	Euryarchaeota	2265081461	A252I15DRAFT_00406	CO dehydrogenase/acetyl-CoA synthase gamma subunit (corrinoid Fe-S protein) (EC:2.1.1.-)	CO dehydrogenase/acetyl-CoA synthase	1371	456
Energy	Reductive acetyl-CoA pathway	55	SCGC_AAA252-I15	Euryarchaeota	2265081460	A252I15DRAFT_00405	CO dehydrogenase/acetyl-CoA synthase, delta subunit	CO dehydrogenase/acetyl-CoA synthase	1170	389
	Reductive acetyl-CoA pathway	55	SCGC_AAA252-I15	Euryarchaeota	2265081457	A252I15DRAFT_00402	CO dehydrogenase/CO-methylating acetyl-CoA synthase complex, beta subunit (EC:2.3.1.-)	CO dehydrogenase/acetyl-CoA synthase	1371	456
	Methane	55	SCGC_AAA252-I15	Euryarchaeota	2265081462	A252I15DRAFT_00407	CO dehydrogenase/acetyl-CoA synthase gamma subunit (corrinoid Fe-S protein) (EC:1.2.99.5)	Methanogenesis by CO2 reduction	447	148

Methane	55	SCGC_AAA252-I15	Euryarchaeota	2265081463	A252I15DRAFT_00408	formylmethanofuran dehydrogenase subunit A (EC:1.2.99.5)	Methanogenesis by CO2 reduction	1590	529
Methane	55	SCGC_AAA252-I15	Euryarchaeota	2265081465	A252I15DRAFT_00410	formylmethanofuran dehydrogenase subunit B (EC:1.2.99.5)	Methanogenesis by CO2 reduction	1308	435
Methane	55	SCGC_AAA252-I15	Euryarchaeota	2265081464	A252I15DRAFT_00409	formylmethanofuran dehydrogenase subunit C (EC:1.2.99.5)	Methanogenesis by CO2 reduction	777	258
Methane	55	SCGC_AAA252-I15	Euryarchaeota	2265081466	A252I15DRAFT_00411	Formylmethanofuran dehydrogenase subunit D (EC:1.2.99.5)	Methanogenesis by CO2 reduction	453	150
Methane	55	SCGC_AAA252-I15	Euryarchaeota	2265081913	A252I15DRAFT_00858	N5-methyltetrahydromethanopterin:coenzyme M methyltransferase subunit H (EC:2.1.1.86)	Methanogenesis by CO2 reduction	831	276
Methane	55	SCGC_AAA252-I15	Euryarchaeota	2265081071	A252I15DRAFT_00016	thymidylate synthase, methanogen type (EC:2.1.1.45)	Methanogenesis by CO2 reduction	633	210
Sulfur	171	JGI_0000059-M03	Caldiserica	2264940980	J0000059M03DRAFT_00012	sulfite reductase, dissimilatory-type alpha subunit (EC:1.8.99.3)	dissimilatory sulfite reductase	1164	387
Sulfur	171	JGI_0000059-M03	Caldiserica	2264940981	J0000059M03DRAFT_00013	sulfite reductase, dissimilatory-type beta subunit (EC:1.8.99.3)	dissimilatory sulfite reductase	1065	354
Sulfur	288	JGI_0000106-J15	Aigarchaeota	2265092802	J0000106J15DRAFT_00083	sulfite reductase, dissimilatory-type alpha subunit (EC:1.8.99.3)	dissimilatory sulfite reductase	1179	392
Sulfur	288	JGI_0000106-J15	Aigarchaeota	2265092803	J0000106J15DRAFT_00084	sulfite reductase, dissimilatory-type beta subunit (EC:1.8.99.3)	dissimilatory sulfite reductase	1074	357
Ammonia	783	SCGC_AAA007-O23	Thaumarchaeota	2506117813	A7O23_0018.00000100	Ammonia monooxygenase/methane monooxygenase, subunit C.	Amonium-oxidizer	564	187
Ammonia	783	SCGC_AAA007-O23	Thaumarchaeota	2506117680	A7O23_0013.00000550	Monooxygenase subunit B protein.	Amonium-oxidizer	603	200
Hydrogen	139	SCGC_AAA257-C11	BRC1	2264938205	A257C11DRAFT_01708	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; 2 subunits, similar to pdb 1CC1/L from Desulfomicrobium baculatum; possibly periplasmic</i>	1512	503
Hydrogen	70	SCGC_AAA252-L23	Aerophobetes	2264944155	A252L23DRAFT_00359	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen sensing; 2 subunits, likely cytosolic; similar to AAX40740.1 from Thiocapsa roseopersicina</i>	1476	491
Hydrogen	70	SCGC_AAA252-L23	Aerophobetes	2264944356	A252L23DRAFT_00561	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound respiratory hydrogenase</i>	1215	404
Hydrogen	129	SCGC_AAA255-F10	Aerophobetes	2264942038	A255F10DRAFT_00506	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound respiratory hydrogenase, similar to AAP51029</i>	1194	397
Hydrogen	281	JGI_0000106-F11	Aenigmarchaea	2264946358	J0000106F11DRAFT_00144	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; 2 subunits</i>	1278	425
Hydrogen	206	SCGC_AAA011-E11	Diapherotrites	2264949134	A11E11DRAFT_00442	<i>Iron only hydrogenase</i>	<i>Hydrogen production; 1 subunit; cytosolic, similar to ACA51661 from Thermoanaerobacterium saccharolyticum JW/SL-YS485</i>	1092	363
Hydrogen	172	JGI_0000059-N10	Hydrogenedentes	2265009594	J0000059N10DRAFT_00090	<i>Iron only hydrogenase</i>	<i>Hydrogen production? multi-subunit, cytosolic</i>	1779	592
Hydrogen	172	JGI_0000059-N10	Hydrogenedentes	2265009559	J0000059N10DRAFT_00055	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1446	481

Hydrogen	236	JGI_J0000077-D07	Hydrogenedentes	2265011014	J0000077D07 DRAFT_00465	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1473	490
Hydrogen	237	JGI_J0000077-O16	Hydrogenedentes	2265013166	J0000077O16 DRAFT_01403	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	435	144
Hydrogen	115	JGI_0000001-D5	Fervidibacteria	2264954473	J0000001D05 DRAFT_00287	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1734	577
Hydrogen	175	JGI_0000054-H05	Fervidibacteria	2264955994	J0000054H05 DRAFT_00197	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Possibly bidirectional; 1 subunit; cytosolic; selenocysteine-containing, similar to BAE46796 from Nostoc sp. PCC 7422</i>	546	181
Hydrogen	178	SCGC_AAA036-E14	Parcubacteria	2265013803	A36E14DRAFT_00490	<i>Iron only hydrogenase</i>	<i>Hydrogen production; 2 subunits; cytosolic, similar to ACA51661 from Thermoanaerobacterium saccharolyticum JW/SL-YS485</i>	1722	573
Hydrogen	240	JGI_0000077-K19	Armatimonadetes	2265018027	J0000077K19 DRAFT_01038	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption?</i>	1509	502
Hydrogen	240	JGI_0000077-K19	Armatimonadetes	2265017046	J0000077K19 DRAFT_00054	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1080	359
Hydrogen	34	SCGC_AAA252-B19	Omnitrophica	2265031127	A252B19DRAFT_00179	<i>Iron only hydrogenase</i>	<i>Hydrogen production; 2 subunits; cytosolic</i>	1899	632
Hydrogen	263	SCGC_AAA011-A17	Omnitrophica	2265030188	A11A17DRAFT_01038	<i>Iron only hydrogenase</i>	<i>Hydrogen production?</i>	1746	581
Hydrogen	126	SCGC_AAA255-E10	Aminicenantes	2265033999	A255E10DRAFT_00411	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit; cytosolic?</i>	1164	387
Hydrogen	60	SCGC_AAA252-K06	Aminicenantes	2265056350	A252K6DRAFT_00369	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit; cytosolic?</i>	1764	587
Hydrogen	232	JGI_0000079-F20	Atribacteria	2265079585	J0000079F20DRAFT_00118	<i>Iron only hydrogenase</i>	<i>Hydrogen production; ? subunits; cytosolic</i>	1473	490
Hydrogen	232	JGI_0000079-F20	Atribacteria	2265079580	J0000079F20DRAFT_00113	<i>Iron only hydrogenase</i>	<i>Hydrogen production; ? subunits; cytosolic; 2 catalytic subunits in 1 chromosomal cluster</i>	1716	571
Hydrogen	231	JGI_0000079-L04	Atribacteria	2265079436	J0000079L04DRAFT_00999	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit; membrane-bound</i>	1224	407
Hydrogen	130	SCGC_AAA255-G05	Atribacteria	2265071497	A255G5DRAFT_00827	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane bound?</i>	1134	377
Hydrogen	136	SCGC_AAA255-N14	Atribacteria	2265073796	A255N14DRAFT_01397	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit?</i>	1329	442
Hydrogen	219	SCGC_AB-164-A22	Atribacteria	2265077615	AB164A22DRAFT_00457	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit; membrane-bound?</i>	1224	407
Hydrogen	218	SCGC_AB-164-M20	Atribacteria	2265076729	AB164M20DRAFT_00076	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit; membrane-bound</i>	1224	407
Hydrogen	215	SCGC_AB-164-P05	Atribacteria	2265075004	AB164P05DRAFT_00184	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit; cytosolic?</i>	1329	442
Hydrogen	55	SCGC_AAA252-I15	Euryarchaeota	2265081183	A252I15DRAFT_00128	<i>Iron only hydrogenase</i>	<i>Hydrogen production; 3 subunits; cytosolic</i>	1725	574
Hydrogen	10	SCGC_AAA471-E14	Aigarchaeota	2265088250	A471E14DRAFT_00499	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; 3 subunits; cytosolic</i>	1041	346
Hydrogen	259	JGI_0000090-C20	Marinimicrobia	2265005968	J0000090C20 DRAFT_00414	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1185	394

Hydrogen	161	JGI_0000039-O11	Marinimicrobia	2265097214	J0000039O11 DRAFT_00168	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit; cytosolic?</i>	1461	486
Hydrogen	161	JGI_0000039-O11	Marinimicrobia	2265097553	J0000039O11 DRAFT_00508	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1185	394
Hydrogen	234	JGI_0000077-B04	Marinimicrobia	2265099776	J0000077B04 DRAFT_00094	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit; cytosolic?</i>	1461	486
Hydrogen	146	SCGC_AAA257-N23	Marinimicrobia	2265095622	A257N23DRAFT_00842	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1191	396
Hydrogen	233	JGI_0000079-D21	Synergistetes	2265107494	J0000079D21 DRAFT_00912	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1239	412
Hydrogen	293	JGI_0000106-O11	Thermotogae	2265108217	J0000106O11 DRAFT_00277	<i>Iron only hydrogenase</i>	<i>Hydrogen production?</i>	1317	438
Hydrogen	293	JGI_0000106-O11	Thermotogae	2265108202	J0000106O11 DRAFT_00262	<i>Iron only hydrogenase</i>	<i>Hydrogen production; 3 subunits; membrane-bound?</i>	2007	668
Hydrogen	168	JGI_0000059-K21	WS1	2265108681	J0000059K21 DRAFT_00371	<i>Iron only hydrogenase</i>	<i>Hydrogen sensing; 1 subunit with PAS domain</i>	1584	527
Hydrogen	145	SCGC_AAA257-K07	Latescibacteria	2265108997	A257K7DRAFT_00275	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1236	411
Hydrogen	42	SCGC_AAA252-E07	Latescibacteria	2265113556	A252E7DRAFT_01033	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1446	481
Hydrogen	42	SCGC_AAA252-E07	Latescibacteria	2265113282	A252E7DRAFT_00759	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1236	411
Hydrogen	32	SCGC_AAA252-B13	Latescibacteria	2265110575	A252B13DRAFT_00290	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multi-subunit membrane-bound</i>	1236	411
Hydrogen	152	JGI_0000039-G13	Cloacimonetes	2265115031	J0000039G13 DRAFT_00184	<i>Iron only hydrogenase</i>	<i>Hydrogen production; multi-subunits; cytosolic?</i>	1785	594
Hydrogen	166	JGI_0000059-F17	Cloacimonetes	2265116390	J0000059F17DRAFT_00292	<i>Iron only hydrogenase</i>	<i>Hydrogen production; 1 subunit; cytosolic</i>	1308	435
Hydrogen	166	JGI_0000059-F17	Cloacimonetes	2265116377	J0000059F17DRAFT_00279	<i>Iron only hydrogenase</i>	<i>Hydrogen production; multi-subunits; cytosolic?</i>	1785	594
Hydrogen	173	JGI_0000059-P03	Cloacimonetes	2265117363	J0000059P03 DRAFT_00329	<i>Iron only hydrogenase</i>	<i>Hydrogen sensing?; multi-subunit; cytosolic?; 2 catalytic subunits clustered with histidine kinase domain protein</i>	1785	594
Hydrogen	173	JGI_0000059-P03	Cloacimonetes	2265117372	J0000059P03 DRAFT_00338	<i>Iron only hydrogenase</i>	<i>Hydrogen sensing?; multi-subunit; cytosolic?; 2 catalytic subunits clustered with histidine kinase domain protein</i>	1215	404
Hydrogen	174	JGI_0000059-P11	Cloacimonetes	2265117635	J0000059P11 DRAFT_00147	<i>Iron only hydrogenase</i>	<i>Hydrogen sensing?; multi-subunit; cytosolic?; 2 catalytic subunits clustered with histidine kinase domain protein</i>	1785	594
Hydrogen	174	JGI_0000059-P11	Cloacimonetes	2265117626	J0000059P11 DRAFT_00138	<i>Iron only hydrogenase</i>	<i>Hydrogen sensing?; multi-subunit; cytosolic?; 2 catalytic subunits clustered with histidine kinase domain protein</i>	1308	435
Hydrogen	274	JGI_0000014-D18	Cloacimonetes	2265118736	J0000014D18 DRAFT_00343	<i>Iron only hydrogenase</i>	<i>Hydrogen production; multi-subunit formate-hydrogen lyase?</i>	1281	426
Hydrogen	273	JGI_0000014_D06	Cloacimonetes	2265118389	J0000014D06 DRAFT_00537	<i>Ni,Fe-hydrogenase large subunit</i>	<i>Hydrogen consumption; multisubunit; cytosolic</i>	1290	429

Hydrogen	37	SCGC_AAA252-D03	Cloacimonetes	2264977298	A252D3DRAFT_00827	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; 2 subunit; cytosolic	1464	487
Hydrogen	37	SCGC_AAA252-D03	Cloacimonetes	2264977331	A252D3DRAFT_00860	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	735	245
Hydrogen	43	SCGC_AAA252-E13	Cloacimonetes	2264978234	A252E13DRAFT_00773	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1218	405
Hydrogen	44	SCGC_AAA252-E20	Cloacimonetes	2264980557	A252E20DRAFT_01485	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; unknown	1464	487
Hydrogen	51	SCGC_AAA252-G08	Cloacimonetes	2264982406	A252G8DRAFT_00535	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1218	405
Hydrogen	53	SCGC_AAA252-I11	Cloacimonetes	2264984835	A252I11DRAFT_01285	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; possibly multi-subunit membrane-bound	1464	487
Hydrogen	53	SCGC_AAA252-I11	Cloacimonetes	2264983962	A252I11DRAFT_00410	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1218	405
Hydrogen	56	SCGC_AAA252-I16	Cloacimonetes	2264986324	A252I16DRAFT_00954	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; possibly multi-subunit membrane-bound	1464	487
Hydrogen	72	SCGC_AAA252-M08	Cloacimonetes	2264989736	A252M8DRAFT_00005	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; possibly multi-subunit membrane-bound	1464	487
Hydrogen	74	SCGC_AAA252-N05	Cloacimonetes	2264991985	A252N5DRAFT_00702	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit; cytosolic?	1464	487
Hydrogen	74	SCGC_AAA252-N05	Cloacimonetes	2264991329	A252N5DRAFT_00046	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1218	405
Hydrogen	77	SCGC_AAA252-N11	Cloacimonetes	2264992322	A252N11DRAFT_00157	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1218	405
Hydrogen	78	SCGC_AAA252-N14	Cloacimonetes	2264994523	A252N14DRAFT_01253	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; possibly multi-subunit membrane-bound	1464	487
Hydrogen	78	SCGC_AAA252-N14	Cloacimonetes	2264994193	A252N14DRAFT_00923	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1125	374
Hydrogen	79	SCGC_AAA252-N17	Cloacimonetes	2264997098	A252N17DRAFT_01812	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; possibly multi-subunit membrane-bound	1464	487
Hydrogen	79	SCGC_AAA252-N17	Cloacimonetes	2264996517	A252N17DRAFT_01230	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1218	405
Hydrogen	87	SCGC_AAA252-O17	Cloacimonetes	2264998514	A252O17DRAFT_00147	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; possibly multi-subunit membrane-bound	1464	487
Hydrogen	87	SCGC_AAA252-O17	Cloacimonetes	2264998596	A252O17DRAFT_00229	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1218	405
Hydrogen	90	SCGC_AAA252-P10	Cloacimonetes	2265001471	A252P10DRAFT_00500	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; possibly multi-subunit membrane-bound	1464	487
Hydrogen	90	SCGC_AAA252-P10	Cloacimonetes	2265002274	A252P10DRAFT_01304	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1218	405
Hydrogen	92	SCGC_AAA252-P14	Cloacimonetes	2265003653	A252P14DRAFT_00721	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; possibly multi-subunit membrane-bound	1464	487
Hydrogen	121	SCGC_AAA255-B16	Cloacimonetes	2264971417	A255B16DRAFT_00126	Ni,Fe-hydrogenase large subunit	Hydrogen consumption; multi-subunit membrane-bound	1218	405
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030615	A11A17DRAFT_01465	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	1083	360
e-chain	268	SCGC_AAA003-E22	Marinimicrobia	2265100580	A03E22DRAFT_00419	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	1116	371

e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101184	A03L08DRAFT_00235	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	1119	372
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506118248	A7O23_0062.00000010	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	228	75
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089953	A471E16DRAFT_00463	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	1047	348
e-chain	171	JGI_0000059-M03	Caldiserica	2264941016	J0000059M03DRAFT_00048	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	1155	384
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030820	A11A17DRAFT_01671	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	1161	386
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090557	A471F17DRAFT_00494	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	1329	442
e-chain	4	SCGC_AAA471-B22	Aigarchaeota	2518715940	A471B22DRAFT_00278	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	1329	442
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117758	A7O23_0014.00000380	Cytochrome b subunit of the bc complex	quinol-cytochrome oxidoreductase	1464	487
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030313	A11A17DRAFT_01163	Cytochrome b subunit of formate dehydrogenase	quinol-cytochrome oxidoreductase	1869	622
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101183	A03L08DRAFT_00234	Cbb3-type cytochrome oxidase, cytochrome c subunit	cytochrome/quinol oxidase-aerobic	2979	992
e-chain	122	SCGC_AAA255-C06	Acetothermia	2265016719	A255C6DRAFT_00502	Uncharacterized anaerobic dehydrogenase	NADH-quinone oxidoreductase	1257	419
e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265110711	A252B13DRAFT_00426	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	231	JGI_0000079-L04	Atribacteria	2265079018	J0000079L04DRAFT_00576	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1197	398
e-chain	232	JGI_0000079-F20	Atribacteria	2265079583	J0000079F20DRAFT_00116	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1899	632
e-chain	293	JGI_0000106-O11	Thermotogae	2265108203	J0000106O11DRAFT_00263	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1905	634
e-chain	55	SCGC_AAA252-I15	Euryarchaeota	2265081184	A252I15DRAFT_00129	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1806	601
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264941874	A255F10DRAFT_00342	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1914	637
e-chain	149	JGI_0000039-D08	Marinimicrobia	2265095744	J0000039D08DRAFT_00052	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G)	NADH-quinone oxidoreductase	849	282
e-chain	60	SCGC_AAA252-K06	Aminicenantes	2265056166	A252K6DRAFT_00185	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3081	1026
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073511	A255N14DRAFT_01110	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3054	1017
e-chain	231	JGI_0000079-L04	Atribacteria	2265079461	J0000079L04DRAFT_01024	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3057	1018
e-chain	48	SCGC_AAA252-G05	Aminicenantes	2265047539	A252G5DRAFT_00894	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	2346	781
e-chain	54	SCGC_AAA252-I14	Aminicenantes	2265050749	A252I14DRAFT_00253	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3153	1050
e-chain	93	SCGC_AAA252-P19	Aminicenantes	2265070370	A252P19DRAFT_01150	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3153	1050

e-chain	84	SCGC_AAA252-O10	Aminicenantes	2265065821	A252O10DRAFT_00160	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3162	1053
e-chain	40	SCGC_AAA252-D18	Aminicenantes	2265044327	A252D18DRAFT_01255	NADPH-dependent glutamate synthase beta chain and related oxidoreductases	NADH-quinone oxidoreductase	1935	644
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958937	A471M6DRAFT_00874	Cbb3-type cytochrome oxidase, subunit 1	cytochrome/quinol oxidase-aerobic	438	145
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265092833	J0000106J15DRAFT_00114	Cytochrome C and Quinol oxidase polypeptide I.	cytochrome/quinol oxidase-aerobic	510	169
e-chain	778	SCGC_AAA298-D23	Marinimicrobia	2265103953	DRAFT_00741	cytochrome c oxidase, subunit I (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1623	540
e-chain	780	SCGC_AAA160-B08	Marinimicrobia	2265105682	DRAFT_00531	cytochrome c oxidase, subunit I (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1623	540
e-chain	291	JGI_0000106-N5	Calescamantes	2264963966	J0000106N5DRAFT_00174	Heme/copper-type cytochrome/quinol oxidases, subunit 1	cytochrome/quinol oxidase-aerobic	1032	343
e-chain	292	JGI_0000106-N7	Calescamantes	2264964909	J00000106N7DRAFT_00509	Heme/copper-type cytochrome/quinol oxidases, subunit 1	cytochrome/quinol oxidase-aerobic	1032	343
e-chain	10	SCGC_AAA471-E14	Aigarchaeota	2265088295	A471E14DRAFT_00544	Heme/copper-type cytochrome/quinol oxidases, subunit 1	cytochrome/quinol oxidase-aerobic	1284	428
e-chain	18	SCGC_AAA471-J07	Aigarchaeota	2265092118	A471J7DRAFT_00258	Heme/copper-type cytochrome/quinol oxidases, subunit 1	cytochrome/quinol oxidase-aerobic	1488	495
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265120973	A471L14DRAFT_00050	Heme/copper-type cytochrome/quinol oxidases, subunit 1	cytochrome/quinol oxidase-aerobic	1626	541
e-chain	291	JGI_0000106-N5	Calescamantes	2264963965	J0000106N5DRAFT_00173	Heme/copper-type cytochrome/quinol oxidases, subunit 1	cytochrome/quinol oxidase-aerobic	732	243
e-chain	292	JGI_0000106-N7	Calescamantes	2264964910	J00000106N7DRAFT_00510	Heme/copper-type cytochrome/quinol oxidases, subunit 1	cytochrome/quinol oxidase-aerobic	732	243
e-chain	112	JGI_0000001-A7	Aigarchaeota	2265088514	J0000001A07DRAFT_00201	Heme/copper-type cytochrome/quinol oxidases, subunit 1	cytochrome/quinol oxidase-aerobic	819	272
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117878	A7023_0002.00000380	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1542	513
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093452	J0000106J15DRAFT_00733	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1590	529
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867064	A471O8DRAFT_00065	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1602	533
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870641	A471C3DRAFT_00143	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1602	533
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868585	A471B5DRAFT_00299	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1605	534
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121041	A471L14DRAFT_00118	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1605	534
e-chain	777	SCGC_AAA160-I06	Marinimicrobia	2265103151	DRAFT_00947	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1623	540
e-chain	781	SCGC_AAA076-M08	Marinimicrobia	2265106552	DRAFT_00406	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1623	540
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264871294	A471C3DRAFT_00796	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1632	543
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868515	A471B5DRAFT_00229	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1650	549

e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869551	A471B23DRAFT_00308	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1650	549
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120521	A471L13DRAFT_00607	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1650	549
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265120943	A471L14DRAFT_00020	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1650	549
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868897	A471B5DRAFT_00611	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1674	557
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264870242	A471B23DRAFT_00999	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1674	557
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870964	A471C3DRAFT_00466	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1674	557
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121512	A471L14DRAFT_00589	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1674	557
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264959526	A471M6DRAFT_01463	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1683	560
e-chain	286	JGI_0000106-I17	Calescamantes	2264961407	J0000106I17DRAFT_00546	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1683	560
e-chain	287	JGI_0000106-I5	Calescamantes	2264961565	J0000106I5DRAFT_00062	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1683	560
e-chain	290	JGI_0000106-M22	Calescamantes	2264962891	J0000106M22DRAFT_00064	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1683	560
e-chain	292	JGI_0000106-N7	Calescamantes	2264965232	J00000106N7DRAFT_00832	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1683	560
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958089	A471M6DRAFT_00026	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1716	571
e-chain	283	JGI_0000106_H18	Calescamantes	2264960650	J0000106H18DRAFT_00322	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1716	571
e-chain	289	JGI_0000106-J16	Calescamantes	2264962707	J0000106J16DRAFT_00890	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1716	571
e-chain	290	JGI_0000106-M22	Calescamantes	2264963733	J0000106M22DRAFT_00906	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1716	571
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265029372	A11A17DRAFT_00222	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1731	576
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101177	A03L08DRAFT_00228	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	1740	579
e-chain	114	JGI_0000001-B9	Fervidibacteria	2264954148	J0000001B09DRAFT_00823	Nitric oxide reductase large subunit (EC:1.7.2.5)	cytochrome/quinol oxidase-aerobic	2136	711
e-chain	277	JGI_0000069_K10	Gracilibacteria	2265148581	J0000069K10DRAFT_00172	Nitric oxide reductase large subunit (EC:1.7.2.5)	cytochrome/quinol oxidase-aerobic	2589	862
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264868223	A471O8DRAFT_01224	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.10.3.12)	cytochrome/quinol oxidase-aerobic	2346	781
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869633	A471B23DRAFT_00390	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.10.3.12)	cytochrome/quinol oxidase-aerobic	2346	781
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870936	A471C3DRAFT_00438	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.10.3.12)	cytochrome/quinol oxidase-aerobic	2346	781
e-chain	118	JGI_0000001-H6	Aigarchaeota	2265089369	J0000001H6DRAFT_00115	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	2334	777

e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090386	A471F17DRAFT_00323	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	2334	777
e-chain	9	SCGC_AAA471-D15	Aigarchaeota	2265094190	A471D15DRAFT_00003	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	2334	777
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089667	A471E16DRAFT_00177	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	2337	778
e-chain	14	SCGC_AAA471-G05	Aigarchaeota	2265091158	A471G5DRAFT_00233	Heme/copper-type cytochrome/quinol oxidases, subunit 1 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	2385	794
e-chain	118	JGI_0000001-H6	Aigarchaeota	2265089368	J0000001H6DRAFT_00114	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	624	207
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089666	A471E16DRAFT_00176	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	624	207
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090385	A471F17DRAFT_00322	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	624	207
e-chain	9	SCGC_AAA471-D15	Aigarchaeota	2265094191	A471D15DRAFT_00004	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	624	207
e-chain	10	SCGC_AAA471-E14	Aigarchaeota	2265088296	A471E14DRAFT_00545	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	636	211
e-chain	14	SCGC_AAA471-G05	Aigarchaeota	2265091157	A471G5DRAFT_00232	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	651	216
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958090	A471M6DRAFT_00027	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	726	241
e-chain	283	JGI_0000106_H18	Calescamantes	2264960651	J0000106H18DRAFT_00323	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	726	241
e-chain	289	JGI_0000106-J16	Calescamantes	2264962706	J0000106J16DRAFT_00889	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	726	241
e-chain	290	JGI_0000106-M22	Calescamantes	2264963732	J0000106M22DRAFT_00905	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	726	241
e-chain	291	JGI_0000106-N5	Calescamantes	2264963964	J0000106N5DRAFT_00172	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	726	241
e-chain	292	JGI_0000106-N7	Calescamantes	2264964911	J00000106N7DRAFT_00511	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	726	241
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264868203	A471O8DRAFT_01204	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	219	72
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869738	A471B23DRAFT_00495	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	219	72
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870870	A471C3DRAFT_00372	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	219	72
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120023	A471L13DRAFT_00109	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	219	72
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121652	A471L14DRAFT_00729	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	219	72
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868416	A471B5DRAFT_00130	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	432	143
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869783	A471B23DRAFT_00540	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	432	143
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870793	A471C3DRAFT_00295	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	432	143

e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121824	A471L14DRAFT_00901	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	432	143
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867805	A471O8DRAFT_00806	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	462	153
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120302	A471L13DRAFT_00388	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	462	153
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121697	A471L14DRAFT_00774	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	462	153
e-chain	55	SCGC_AAA252-I15	Euryarchaeota	2265082429	A252I15DRAFT_01374	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	468	155
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101172	A03L08DRAFT_00223	Heme/copper-type cytochrome/quinol oxidases, subunit 2	cytochrome/quinol oxidase-aerobic	927	308
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264868224	A471O8DRAFT_01225	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.10.3.12)	cytochrome/quinol oxidase-aerobic	426	141
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869634	A471B23DRAFT_00391	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.10.3.12)	cytochrome/quinol oxidase-aerobic	426	141
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870937	A471C3DRAFT_00439	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.10.3.12)	cytochrome/quinol oxidase-aerobic	426	141
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265120974	A471L14DRAFT_00051	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.10.3.12)	cytochrome/quinol oxidase-aerobic	426	141
e-chain	205	SCGC_AAA011-L05	Microgenomates	2265023496	A11L05DRAFT_00186	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	390	129
e-chain	199	SCGC_AAA011-K09	Diapherotrites	2264948348	A11K09DRAFT_00040	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	411	136
e-chain	193	SCGC_AAA011-A19	Microgenomates	2265021022	A11A19DRAFT_00036	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	417	138
e-chain	192	SCGC_AAA011-L22	Nanoarchaeota	2264950784	A11L22DRAFT_00458	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	423	140
e-chain	198	SCGC_AAA011-G17	Nanoarchaeota	2264951533	A11G17DRAFT_00712	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	423	140
e-chain	202	SCGC_AAA011-K22	Nanoarchaeota	2264952612	A11K22DRAFT_00423	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	423	140
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117879	A7023_0002.00000390	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	423	140
e-chain	191	SCGC_AAA011-J21	Parcubacteria	2265015169	A11J21DRAFT_00263	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	432	143
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867063	A471O8DRAFT_00064	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	456	151
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868584	A471B5DRAFT_00298	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	456	151
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870640	A471C3DRAFT_00142	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	456	151
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121042	A471L14DRAFT_00119	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	456	151
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868514	A471B5DRAFT_00228	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	519	172
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868898	A471B5DRAFT_00612	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	519	172

e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869552	A471B23DRAFT_00309	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	519	172
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264870241	A471B23DRAFT_00998	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	519	172
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870963	A471C3DRAFT_00465	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	519	172
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120522	A471L13DRAFT_00608	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	519	172
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265120942	A471L14DRAFT_00019	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	519	172
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121511	A471L14DRAFT_00588	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	519	172
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264959525	A471M6DRAFT_01462	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	528	175
e-chain	286	JGI_0000106-I17	Calescamantes	2264961408	J0000106I17DRAFT_00547	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	528	175
e-chain	290	JGI_0000106-M22	Calescamantes	2264962890	J0000106M22DRAFT_00063	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	528	175
e-chain	292	JGI_0000106-N7	Calescamantes	2264965233	J00000106N7DRAFT_00833	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	528	175
e-chain	112	JGI_0000001-A7	Aigarchaeota	2265088507	J0000001A07DRAFT_00194	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	546	181
e-chain	18	SCGC_AAA471-J07	Aigarchaeota	2265092117	A471J7DRAFT_00257	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	549	182
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264871293	A471C3DRAFT_00795	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	552	183
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093453	J0000106J15DRAFT_00734	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	573	190
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265029371	A11A17DRAFT_00221	Heme/copper-type cytochrome/quinol oxidases, subunit 2 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	783	260
e-chain	14	SCGC_AAA471-G05	Aigarchaeota	2265091406	A471G5DRAFT_00481	Nitrous oxide reductase	cytochrome/quinol oxidase-aerobic	321	106
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958703	A471M6DRAFT_00640	Nitrous oxide reductase (EC:1.7.2.4)	cytochrome/quinol oxidase-aerobic	1995	664
e-chain	282	JGI_0000106_G12	Calescamantes	2264960039	J0000106G12DRAFT_00224	Nitrous oxide reductase (EC:1.7.2.4)	cytochrome/quinol oxidase-aerobic	1995	664
e-chain	286	JGI_0000106-I17	Calescamantes	2264960960	J0000106I17DRAFT_00098	Nitrous oxide reductase (EC:1.7.2.4)	cytochrome/quinol oxidase-aerobic	1995	664
e-chain	289	JGI_0000106-J16	Calescamantes	2264962385	J0000106J16DRAFT_00568	Nitrous oxide reductase (EC:1.7.2.4)	cytochrome/quinol oxidase-aerobic	1995	664
e-chain	290	JGI_0000106-M22	Calescamantes	2264962934	J0000106M22DRAFT_00107	Nitrous oxide reductase (EC:1.7.2.4)	cytochrome/quinol oxidase-aerobic	1995	664
e-chain	292	JGI_0000106-N7	Calescamantes	2264964417	J00000106N7DRAFT_00017	Nitrous oxide reductase (EC:1.7.2.4)	cytochrome/quinol oxidase-aerobic	1995	664
e-chain	294	JGI_0000106-P5	Calescamantes	2264966037	J0000106P5DRAFT_00276	Nitrous oxide reductase (EC:1.7.2.4)	cytochrome/quinol oxidase-aerobic	1995	664
e-chain	19	SCGC_AAA471-J08	Aigarchaeota	2265092224	A471J8DRAFT_00066	Nitrous oxide reductase (EC:1.7.2.4)	cytochrome/quinol oxidase-aerobic	2052	683

e-chain	777	SCGC_AAA160-I06	Marinimicrobia	2265103152	DRAFT_00948	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	972	323
e-chain	780	SCGC_AAA160-B08	Marinimicrobia	2265105683	DRAFT_00532	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	972	323
e-chain	781	SCGC_AAA076-M08	Marinimicrobia	2265106553	DRAFT_00407	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	972	323
e-chain	778	SCGC_AAA298-D23	Marinimicrobia	2265103952	DRAFT_00740	cytochrome c oxidase, subunit II (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	975	324
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073773	A255N14DRAFT_01374	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1152	383
e-chain	122	SCGC_AAA255-C06	Acetothermia	2265016726	A255C6DRAFT_00509	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1257	418
e-chain	37	SCGC_AAA252-D03	Cloacimonetes	2264977284	A252D3DRAFT_00813	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1275	425
e-chain	130	SCGC_AAA255-G05	Atribacteria	2265071504	A255G5DRAFT_00834	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1425	474
e-chain	78	SCGC_AAA252-N14	Cloacimonetes	2518602968	A252N14PER MD_00876	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1437	478
e-chain	37	SCGC_AAA252-D03	Cloacimonetes	2264977285	A252D3DRAFT_00814	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1470	489
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944678	A252L23DRAFT_00883	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1476	491
e-chain	122	SCGC_AAA255-C06	Acetothermia	2265016825	A255C6DRAFT_00608	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1476	491
e-chain	259	JGI_0000090-C20	Marinimicrobia	2265005973	J0000090C20 DRAFT_00419	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1485	494
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073291	A255N14DRAFT_00890	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1485	494
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077959	AB164A22DR AFT_00807	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1485	494
e-chain	231	JGI_0000079-L04	Atribacteria	2265079432	J0000079L04DRAFT_00995	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1485	494
e-chain	161	JGI_0000039-O11	Marinimicrobia	2265097558	J0000039O11 DRAFT_00513	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1485	494
e-chain	172	JGI_0000059-N10	Hydrogenedentes	2265009990	J0000059N10 DRAFT_00486	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1491	496
e-chain	237	JGI_J00000077-O16	Hydrogenedentes	2265012137	J0000077O16 DRAFT_00372	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1491	496

e-chain	145	SCGC_AAA257-K07	Latescibacteria	2265109002	A257K7DRAFT_00280	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1497	498
e-chain	121	SCGC_AAA255-B16	Cloacimonetes	2264971411	A255B16DRAFT_00120	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	37	SCGC_AAA252-D03	Cloacimonetes	2264977259	A252D3DRAFT_00788	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	51	SCGC_AAA252-G08	Cloacimonetes	2264982400	A252G8DRAFT_00529	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264983968	A252I11DRAFT_00416	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	74	SCGC_AAA252-N05	Cloacimonetes	2264991323	A252N5DRAFT_00040	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	77	SCGC_AAA252-N11	Cloacimonetes	2264992316	A252N11DRAFT_00151	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	79	SCGC_AAA252-N17	Cloacimonetes	2264995839	A252N17DRAFT_00552	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	87	SCGC_AAA252-O17	Cloacimonetes	2264998602	A252O17DRAFT_00235	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	90	SCGC_AAA252-P10	Cloacimonetes	2265002280	A252P10DRAFT_01310	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265110570	A252B13DRAFT_00285	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1506	501
e-chain	218	SCGC_AB-164-M20	Atribacteria	2265076725	AB164M20DRAFT_00072	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1566	521
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077619	AB164A22DRAFT_00461	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1566	521
e-chain	233	JGI_0000079-D21	Synergistetes	2265107490	J0000079D21DRAFT_00908	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1581	526
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944408	A252L23DRAFT_00613	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1701	566
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073290	A255N14DRAFT_00889	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	1788	595
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265113277	A252E7DRAFT_00754	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	579	192
e-chain	74	SCGC_AAA252-N05	Cloacimonetes	2264992105	A252N5DRAFT_00822	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	873	291

e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265113276	A252E7DRAFT_00753	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	NADH-quinone oxidoreductase	900	299
e-chain	115	JGI_0000001-D5	Fervidibacteria	2264954472	J0000001D05 DRAFT_00286	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit (EC:1.-)	NADH-quinone oxidoreductase	1986	661
e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265111381	A252B13DRAFT_01097	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1737	578
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944679	A252L23DRAFT_00884	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1860	619
e-chain	122	SCGC_AAA255-C06	Acetothermia	2265016826	A255C6DRAFT_00609	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1860	619
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506118221	A7O23_0052.00000010	NADH:ubiquinone oxidoreductase subunit 2 (chain N)	NADH-quinone oxidoreductase	318	105
e-chain	48	SCGC_AAA252-G05	Aminicenantes	2265047692	A252G5DRAFT_01047	NADH:ubiquinone oxidoreductase subunit 2 (chain N)	NADH-quinone oxidoreductase	450	149
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506118205	A7O23_0044.00000020	NADH:ubiquinone oxidoreductase subunit 2 (chain N)	NADH-quinone oxidoreductase	513	170
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117335	A7O23_0001.00000960	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1305	434
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958556	A471M6DRAFT_00493	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1404	467
e-chain	286	JGI_0000106-I17	Calescamantes	2264961288	J0000106I17DRAFT_00427	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1404	467
e-chain	287	JGI_0000106-I5	Calescamantes	2264961583	J0000106I5DRAFT_00080	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1404	467
e-chain	290	JGI_0000106-M22	Calescamantes	2264963323	J0000106M22 DRAFT_00496	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1404	467
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867704	A471O8DRAFT_00705	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1455	484
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868342	A471B5DRAFT_00056	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1455	484
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869455	A471B23DRAFT_00212	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1455	484
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120127	A471L13DRAFT_00213	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1455	484
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121269	A471L14DRAFT_00346	NADH:ubiquinone oxidoreductase subunit 2 (chain N) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1455	484
e-chain	259	JGI_0000090-C20	Marinimicrobia	2265005690	J0000090C20 DRAFT_00135	NADH:ubiquinone oxidoreductase subunit 4 (chain M)	NADH-quinone oxidoreductase	1464	487
e-chain	149	JGI_0000039-D08	Marinimicrobia	2265095759	J0000039D08 DRAFT_00067	NADH:ubiquinone oxidoreductase subunit 4 (chain M)	NADH-quinone oxidoreductase	1464	487
e-chain	161	JGI_0000039-O11	Marinimicrobia	2265097256	J0000039O11 DRAFT_00210	NADH:ubiquinone oxidoreductase subunit 4 (chain M)	NADH-quinone oxidoreductase	1464	487
e-chain	169	_JGI_0000059-L03	Marinimicrobia	2265099619	J0000059L03DRAFT_00522	NADH:ubiquinone oxidoreductase subunit 4 (chain M)	NADH-quinone oxidoreductase	1464	487
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117338	A7O23_0001.00000990	NADH:ubiquinone oxidoreductase subunit 4 (chain M) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1551	516

e-chain	146	SCGC_AAA257-N23	Marinimicrobia	2265095617	A257N23DRAFT_00837	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit	NADH-quinone oxidoreductase	1188	395
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264941560	A255F10DRAFT_00026	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit	NADH-quinone oxidoreductase	1710	569
e-chain	66	SCGC_AAA252-L05	Cloacimonetes	2264988868	A252L5DRAFT_00665	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit	NADH-quinone oxidoreductase	738	245
e-chain	115	JGI_0000001-D5	Fervidibacteria	2264954477	J0000001D05DRAFT_00291	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.-)	NADH-quinone oxidoreductase	1455	484
e-chain	115	JGI_0000001-D5	Fervidibacteria	2264954471	J0000001D05DRAFT_00285	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.-)	NADH-quinone oxidoreductase	1491	496
e-chain	78	SCGC_AAA252-N14	Cloacimonetes	2518602973	A252N14PERMD_00881	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1410	469
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264942036	A255F10DRAFT_00504	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1422	473
e-chain	79	SCGC_AAA252-N17	Cloacimonetes	2264996516	A252N17DRAFT_01229	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1566	521
e-chain	121	SCGC_AAA255-B16	Cloacimonetes	2264971416	A255B16DRAFT_00125	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1815	604
e-chain	37	SCGC_AAA252-D03	Cloacimonetes	2264977254	A252D3DRAFT_00783	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1815	604
e-chain	43	SCGC_AAA252-E13	Cloacimonetes	2264978233	A252E13DRAFT_00772	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1815	604
e-chain	51	SCGC_AAA252-G08	Cloacimonetes	2264982405	A252G8DRAFT_00534	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1815	604
e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264983963	A252I11DRAFT_00411	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1815	604
e-chain	74	SCGC_AAA252-N05	Cloacimonetes	2264991328	A252N5DRAFT_00045	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1815	604
e-chain	77	SCGC_AAA252-N11	Cloacimonetes	2264992321	A252N11DRAFT_00156	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1815	604
e-chain	87	SCGC_AAA252-O17	Cloacimonetes	2264998597	A252O17DRAFT_00230	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1815	604
e-chain	90	SCGC_AAA252-P10	Cloacimonetes	2265002275	A252P10DRAFT_01305	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1815	604
e-chain	259	JGI_0000090-C20	Marinimicrobia	2265005969	J0000090C20DRAFT_00415	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1824	607

e-chain	161	JGI_0000039-O11	Marinimicrobia	2265097554	J0000039O11 DRAFT_00509	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1824	607
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944354	A252L23DRAF T_00559	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1926	641
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017050	J0000077K19 DRAFT_00058	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1986	661
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121929	A471L14DRAF T_01006	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1449	482
e-chain	113	JGI_0000001-B8	Aigarchaeota	2265089054	J0000001B8D RAFT_00250	proton-translocating NADH-quinone oxidoreductase, chain M	NADH-quinone oxidoreductase	1467	488
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090580	A471F17DRAF T_00517	proton-translocating NADH-quinone oxidoreductase, chain M	NADH-quinone oxidoreductase	1467	488
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264942031	A255F10DRAF T_00499	proton-translocating NADH-quinone oxidoreductase, chain M	NADH-quinone oxidoreductase	1482	493
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944350	A252L23DRAF T_00555	proton-translocating NADH-quinone oxidoreductase, chain M	NADH-quinone oxidoreductase	1482	493
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093818	J0000106J15D RAFT_01099	proton-translocating NADH-quinone oxidoreductase, chain M	NADH-quinone oxidoreductase	1497	498
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264870301	A471B23DRAF T_01058	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1194	397
e-chain	287	JGI_0000106-I5	Calescamantes	2264961584	J0000106I5DR AFT_00081	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1233	410
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958555	A471M6DRAF T_00492	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1458	485
e-chain	286	JGI_0000106-I17	Calescamantes	2264961287	J0000106I17D RAFT_00426	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1458	485
e-chain	290	JGI_0000106-M22	Calescamantes	2264963324	J0000106M22 DRAFT_00497	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1458	485
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867832	A471O8DRAFT _00833	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1464	487
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120329	A471L13DRAF T_00415	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1464	487
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121928	A471L14DRAF T_01005	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1464	487
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030247	A11A17DRAFT _01097	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1473	490
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265029269	A11A17DRAFT _00119	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1488	495
e-chain	114	JGI_0000001-B9	Fervidibacteria	2264953705	J0000001B09 DRAFT_00379	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1533	510
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101558	A03L08DRAFT _00609	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1572	523
e-chain	60	SCGC_AAA252-K06	Aminicenantes	2265056441	A252K6DRAFT _00460	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1587	528
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017700	J0000077K19 DRAFT_00710	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1614	537

e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030246	A11A17DRAFT_01096	proton-translocating NADH-quinone oxidoreductase, chain N (EC:1.6.5.3)	NADH-quinone oxidoreductase	1419	472
e-chain	226	JGI_0000014-B08	Aminicenantes	2265038112	J0000014B08 DRAFT_00282	proton-translocating NADH-quinone oxidoreductase, chain N (EC:1.6.5.3)	NADH-quinone oxidoreductase	1428	475
e-chain	60	SCGC_AAA252-K06	Aminicenantes	2265056442	A252K6DRAFT_00461	proton-translocating NADH-quinone oxidoreductase, chain N (EC:1.6.5.3)	NADH-quinone oxidoreductase	1434	477
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265029268	A11A17DRAFT_00118	proton-translocating NADH-quinone oxidoreductase, chain N (EC:1.6.5.3)	NADH-quinone oxidoreductase	1443	480
e-chain	117	JGI_0000001-G10	Fervidibacteria	2264955041	J0000001G10 DRAFT_00293	proton-translocating NADH-quinone oxidoreductase, chain N (EC:1.6.5.3)	NADH-quinone oxidoreductase	1449	482
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017457	J0000077K19 DRAFT_00467	proton-translocating NADH-quinone oxidoreductase, chain N (EC:1.6.5.3)	NADH-quinone oxidoreductase	1476	491
e-chain	117	JGI_0000001-G10	Fervidibacteria	2264954932	J0000001G10 DRAFT_00184	proton-translocating NADH-quinone oxidoreductase, chain N (EC:1.6.5.3)	NADH-quinone oxidoreductase	1563	520
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101559	A03L08DRAFT_00610	proton-translocating NADH-quinone oxidoreductase, chain N (EC:1.6.5.3)	NADH-quinone oxidoreductase	1707	568
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117337	A7O23_0001.00000980	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	2073	690
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264942037	A255F10DRAFT_00505	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1416	471
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944355	A252L23DRAFT_00560	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1416	471
e-chain	145	SCGC_AAA257-K07	Latescibacteria	2265108998	A257K7DRAFT_00276	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1836	611
e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265110574	A252B13DRAFT_00289	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1836	611
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265113281	A252E7DRAFT_00758	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit (EC:1.6.99.5)	NADH-quinone oxidoreductase	1839	612
e-chain	287	JGI_0000106-I5	Calescamantes	2264961568	J0000106I5DRAFT_00065	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1602	534
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958554	A471M6DRAFT_00491	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1812	603
e-chain	286	JGI_0000106-I17	Calescamantes	2264961286	J0000106I17DRAFT_00425	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1812	603
e-chain	290	JGI_0000106-M22	Calescamantes	2264963325	J0000106M22 DRAFT_00498	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1812	603
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265029270	A11A17DRAFT_00120	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1842	613
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867831	A471O8DRAFT_00832	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1866	621
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120328	A471L13DRAFT_00414	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1866	621
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264870308	A471B23DRAFT_01065	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1878	625

e-chain	112	JGI_0000001-A7	Aigarchaeota	2265088603	J0000001A07 DRAFT_00290	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1905	634
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017699	J0000077K19 DRAFT_00709	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	2013	670
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101557	A03L08DRAFT _00608	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	2109	702
e-chain	19	SCGC_AAA471-J08	Aigarchaeota	2265092214	A471J8DRAFT _00056	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1452	483
e-chain	226	JGI_0000014-B08	Aminicenantes	2265038110	J0000014B08 DRAFT_00280	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1902	633
e-chain	60	SCGC_AAA252-K06	Aminicenantes	2265056440	A252K6DRAFT _00459	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1908	635
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030248	A11A17DRAFT _01098	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	1917	638
e-chain	114	JGI_0000001-B9	Fervidibacteria	2264953701	J0000001B09 DRAFT_00375	proton-translocating NADH-quinone oxidoreductase, chain L (EC:1.6.5.3)	NADH-quinone oxidoreductase	2010	669
e-chain	117	JGI_0000001-G10	Fervidibacteria	2264955406	J0000001G10 DRAFT_00659	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1281	426
e-chain	226	JGI_0000014-B08	Aminicenantes	2265038111	J0000014B08 DRAFT_00281	proton-translocating NADH-quinone oxidoreductase, chain M (EC:1.6.5.3)	NADH-quinone oxidoreductase	1590	529
e-chain	172	JGI_0000059-N10	Hydrogenedentes	2265009993	J0000059N10 DRAFT_00489	NADH:ubiquinone oxidoreductase subunit 5 (chain L)/Multisubunit Na ⁺ /H ⁺ antiporter, MnhA subunit	NADH-quinone oxidoreductase	2319	772
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944607	A252L23DRAF T_00812	formate dehydrogenase, alpha subunit, archaeal-type (EC:1.2.1.43)	NADH-quinone oxidoreductase	2712	903
e-chain	130	SCGC_AAA255-G05	Atribacteria	2265072141	A255G5DRAFT _01475	formate dehydrogenase, alpha subunit, archaeal-type (EC:1.2.1.43)	NADH-quinone oxidoreductase	2733	910
e-chain	274	JGI_0000014-D18	Cloacimonetes	2265118733	J0000014D18 DRAFT_00340	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G)	NADH-quinone oxidoreductase	2337	778
e-chain	126	SCGC_AAA255-E10	Aminicenantes	2265033846	A255E10DRAF T_00258	Uncharacterized anaerobic dehydrogenase (EC:1.2.1.2)	NADH-quinone oxidoreductase	2454	817
e-chain	231	JGI_0000079-L04	Atribacteria	2265079457	J0000079L04D RAFT_01020	Uncharacterized anaerobic dehydrogenase (EC:1.2.1.43)	NADH-quinone oxidoreductase	2700	899
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077614	AB164A22DR AFT_00456	formate dehydrogenase, alpha subunit, archaeal-type (EC:1.2.1.43)	NADH-quinone oxidoreductase	2610	869
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073514	A255N14DRAF T_01113	Uncharacterized anaerobic dehydrogenase	NADH-quinone oxidoreductase	1647	549
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030536	A11A17DRAFT _01386	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G) (EC:1.6.5.3)	NADH-quinone oxidoreductase	1608	535
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017695	J0000077K19 DRAFT_00705	NADH-quinone oxidoreductase, chain G (EC:1.6.5.3)	NADH-quinone oxidoreductase	2223	740
e-chain	115	JGI_0000001-D5	Fervidibacteria	2264954476	J0000001D05 DRAFT_00290	Hydrogenase 4 membrane component (E) (EC:1.-)	NADH-quinone oxidoreductase	663	220
e-chain	237	JGI_J00000077-O16	Hydrogenedentes	2265012136	J0000077O16 DRAFT_00371	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	237	78
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264941561	A255F10DRAF T_00027	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264942032	A255F10DRAF T_00500	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111

e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944409	A252L23DRAFT_00614	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	121	SCGC_AAA255-B16	Cloacimonetes	2264971412	A255B16DRAFT_00121	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	37	SCGC_AAA252-D03	Cloacimonetes	2264977258	A252D3DRAFT_00787	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	51	SCGC_AAA252-G08	Cloacimonetes	2264982401	A252G8DRAFT_00530	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264983967	A252I11DRAFT_00415	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	74	SCGC_AAA252-N05	Cloacimonetes	2264991324	A252N5DRAFT_00041	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	77	SCGC_AAA252-N11	Cloacimonetes	2264992317	A252N11DRAFT_00152	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	79	SCGC_AAA252-N17	Cloacimonetes	2264995840	A252N17DRAFT_00553	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	87	SCGC_AAA252-O17	Cloacimonetes	2264998601	A252O17DRAFT_00234	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	90	SCGC_AAA252-P10	Cloacimonetes	2265002279	A252P10DRAFT_01309	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	78	SCGC_AAA252-N14	Cloacimonetes	2518602969	A252N14PERMD_00877	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	336	111
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944351	A252L23DRAFT_00556	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	339	112
e-chain	146	SCGC_AAA257-N23	Marinimicrobia	2265095620	A257N23DRAFT_00840	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	339	112
e-chain	259	JGI_0000090-C20	Marinimicrobia	2265005972	J0000090C20DRAFT_00418	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	345	114
e-chain	161	JGI_0000039-O11	Marinimicrobia	2265097557	J0000039O11DRAFT_00512	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	345	114
e-chain	172	JGI_0000059-N10	Hydrogenedentes	2265009991	J0000059N10DRAFT_00487	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	348	115
e-chain	130	SCGC_AAA255-G05	Atribacteria	2265072253	A255G5DRAFT_01590	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	348	115
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073292	A255N14DRAFT_00891	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	348	115
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073772	A255N14DRAFT_01373	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	348	115
e-chain	218	SCGC_AB-164-M20	Atribacteria	2265076724	AB164M20DRAFT_00071	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	348	115
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077620	AB164A22DRAFT_00462	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	348	115
e-chain	231	JGI_0000079-L04	Atribacteria	2265079261	J0000079L04DRAFT_00821	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	348	115
e-chain	233	JGI_0000079-D21	Synergistetes	2265107489	J0000079D21DRAFT_00907	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	351	116
e-chain	145	SCGC_AAA257-K07	Latescibacteria	2265109001	A257K7DRAFT_00279	Multisubunit Na+/H+ antiporter, MnhC subunit	NADH-quinone oxidoreductase	351	116

e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265110571	A252B13DRAFT_00286	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	351	116
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265113278	A252E7DRAFT_00755	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	351	116
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944677	A252L23DRAFT_00882	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	360	119
e-chain	37	SCGC_AAA252-D03	Cloacimonetes	2264977286	A252D3DRAFT_00815	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	360	119
e-chain	74	SCGC_AAA252-N05	Cloacimonetes	2264992106	A252N5DRAFT_00823	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	360	119
e-chain	122	SCGC_AAA255-C06	Acetothermia	2265016824	A255C6DRAFT_00607	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	360	119
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077958	AB164A22DRAFT_00806	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	360	119
e-chain	122	SCGC_AAA255-C06	Acetothermia	2265016795	A255C6DRAFT_00578	Multisubunit Na ⁺ /H ⁺ antiporter, MnhC subunit	NADH-quinone oxidoreductase	363	120
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958553	A471M6DRAFT_00490	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	300	99
e-chain	286	JGI_0000106-I17	Calescamantes	2264961285	J0000106I17DRAFT_00424	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	300	99
e-chain	287	JGI_0000106-I5	Calescamantes	2264961569	J0000106I5DRAFT_00066	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	300	99
e-chain	290	JGI_0000106-M22	Calescamantes	2264963326	J0000106M22DRAFT_00499	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	300	99
e-chain	10	SCGC_AAA471-E14	Aigarchaeota	2265088195	A471E14DRAFT_00444	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	300	99
e-chain	112	JGI_0000001-A7	Aigarchaeota	2265088459	J0000001A07DRAFT_00146	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	300	99
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867703	A471O8DRAFT_00704	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	303	100
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868341	A471B5DRAFT_00055	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	303	100
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869454	A471B23DRAFT_00211	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	303	100
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265029271	A11A17DRAFT_00121	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	303	100
e-chain	226	JGI_0000014-B08	Aminicenantes	2265038109	J0000014B08DRAFT_00279	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	303	100
e-chain	60	SCGC_AAA252-K06	Aminicenantes	2265056439	A252K6DRAFT_00458	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	303	100
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120126	A471L13DRAFT_00212	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	303	100
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121270	A471L14DRAFT_00347	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	303	100
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117339	A7O23_0001.00001000	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	306	101
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101556	A03L08DRAFT_00607	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	309	102

e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017698	J0000077K19 DRAFT_00708	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	312	103
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093824	J0000106J15D RAFT_01105	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	312	103
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030249	A11A17DRAFT _01099	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	315	104
e-chain	9	SCGC_AAA471-D15	Aigarchaeota	2265094233	A471D15DRAF T_00046	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	315	104
e-chain	114	JGI_0000001-B9	Fervidibacteria	2264953696	J0000001B09 DRAFT_00370	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	318	105
e-chain	113	JGI_0000001-B8	Aigarchaeota	2265088836	J0000001B8D RAFT_00032	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	321	106
e-chain	118	JGI_0000001-H6	Aigarchaeota	2265089349	J0000001H6D RAFT_00095	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	321	106
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089831	A471E16DRAF T_00341	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	321	106
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090707	A471F17DRAF T_00644	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	321	106
e-chain	14	SCGC_AAA471-G05	Aigarchaeota	2265091236	A471G5DRAFT _00311	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	321	106
e-chain	4	SCGC_AAA471-B22	Aigarchaeota	2518715723	A471B22DRAF T_00061	NADH:ubiquinone oxidoreductase subunit 11 or 4L (chain K) (EC:1.6.5.3)	NADH-quinone oxidoreductase	321	106
e-chain	10	SCGC_AAA471-E14	Aigarchaeota	2265088193	A471E14DRAF T_00442	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	249	82
e-chain	112	JGI_0000001-A7	Aigarchaeota	2265088461	J0000001A07 DRAFT_00148	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	249	82
e-chain	113	JGI_0000001-B8	Aigarchaeota	2265088834	J0000001B8D RAFT_00030	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	252	83
e-chain	118	JGI_0000001-H6	Aigarchaeota	2265089347	J0000001H6D RAFT_00093	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	252	83
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089829	A471E16DRAF T_00339	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	252	83
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090705	A471F17DRAF T_00642	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	252	83
e-chain	14	SCGC_AAA471-G05	Aigarchaeota	2265091238	A471G5DRAFT _00313	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	252	83
e-chain	9	SCGC_AAA471-D15	Aigarchaeota	2265094235	A471D15DRAF T_00048	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	252	83
e-chain	4	SCGC_AAA471-B22	Aigarchaeota	2518715721	A471B22DRAF T_00059	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	252	83
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093826	J0000106J15D RAFT_01107	NADH-ubiquinone/plastoquinone oxidoreductase chain 6.	NADH-quinone oxidoreductase	273	90
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867702	A471O8DRAFT _00703	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	483	160
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868340	A471B5DRAFT _00054	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	483	160
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869453	A471B23DRAF T_00210	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	483	160

e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120125	A471L13DRAFT_00211	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	483	160
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121271	A471L14DRAFT_00348	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	483	160
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101555	A03L08DRAFT_00606	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	495	164
e-chain	60	SCGC_AAA252-K06	Aminicenantes	2265056438	A252K6DRAFT_00457	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	498	165
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030250	A11A17DRAFT_01100	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	510	169
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117340	A7O23_0001.00001010	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	516	171
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958552	A471M6DRAFT_00489	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	519	172
e-chain	286	JGI_0000106-I17	Calescamantes	2264961284	J0000106I17DRAFT_00423	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	519	172
e-chain	287	JGI_0000106-I5	Calescamantes	2264961570	J0000106I5DRAFT_00067	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	519	172
e-chain	290	JGI_0000106-M22	Calescamantes	2264963327	J0000106M22DRAFT_00500	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	519	172
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017697	J0000077K19DRAFT_00707	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	519	172
e-chain	114	JGI_0000001-B9	Fervidibacteria	2264953695	J0000001B09DRAFT_00369	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	522	173
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265029272	A11A17DRAFT_00122	NADH:ubiquinone oxidoreductase subunit 6 (chain J) (EC:1.6.5.3)	NADH-quinone oxidoreductase	561	186
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030531	A11A17DRAFT_01381	NADH:ubiquinone oxidoreductase subunit 3 (chain A)	NADH-quinone oxidoreductase	513	170
e-chain	126	SCGC_AAA255-E10	Aminicenantes	2265034247	A255E10DRAFT_00659	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	312	104
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867696	A471O8DRAFT_00697	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	357	118
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868334	A471B5DRAFT_00048	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	357	118
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869447	A471B23DRAFT_00204	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	357	118
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120119	A471L13DRAFT_00205	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	357	118
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121277	A471L14DRAFT_00354	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	357	118
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030256	A11A17DRAFT_01106	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	366	121
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117346	A7O23_0001.00001070	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	375	124
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264959691	A471M6DRAFT_01629	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	381	126
e-chain	283	JGI_0000106_H18	Calescamantes	2264960386	J0000106H18DRAFT_00058	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	381	126

e-chain	289	JGI_0000106-J16	Calescamantes	2264962181	J0000106J16DRAFT_00364	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	381	126
e-chain	290	JGI_0000106-M22	Calescamantes	2264963079	J0000106M22DRAFT_00252	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	381	126
e-chain	292	JGI_0000106-N7	Calescamantes	2264965511	J00000106N7DRAFT_01111	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	381	126
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101135	A03L08DRAFT_00186	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	417	138
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017690	J0000077K19DRAFT_00700	NADH:ubiquinone oxidoreductase subunit 3 (chain A) (EC:1.6.5.3)	NADH-quinone oxidoreductase	489	162
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958088	A471M6DRAFT_00025	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	588	195
e-chain	283	JGI_0000106_H18	Calescamantes	2264960649	J0000106H18DRAFT_00321	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	588	195
e-chain	289	JGI_0000106-J16	Calescamantes	2264962708	J0000106J16DRAFT_00891	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	588	195
e-chain	290	JGI_0000106-M22	Calescamantes	2264963734	J0000106M22DRAFT_00907	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	588	195
e-chain	291	JGI_0000106-N5	Calescamantes	2264963967	J0000106N5DRAFT_00175	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	588	195
e-chain	292	JGI_0000106-N7	Calescamantes	2264964908	J00000106N7DRAFT_00508	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	588	195
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101176	A03L08DRAFT_00227	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	591	196
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265029373	A11A17DRAFT_00223	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	600	199
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958979	A471M6DRAFT_00916	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	633	210
e-chain	292	JGI_0000106-N7	Calescamantes	2264965142	J00000106N7DRAFT_00742	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	633	210
e-chain	777	SCGC_AAA160-I06	Marinimicrobia	2265103150	DRAFT_00946	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	654	217
e-chain	780	SCGC_AAA160-B08	Marinimicrobia	2265105681	DRAFT_00530	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	654	217
e-chain	781	SCGC_AAA076-M08	Marinimicrobia	2265106551	DRAFT_00405	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	654	217
e-chain	778	SCGC_AAA298-D23	Marinimicrobia	2265103954	DRAFT_00742	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	660	219
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265120972	A471L14DRAFT_00049	Heme/copper-type cytochrome/quinol oxidase, subunit 3 (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	732	243
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090911	A471F17DRAFT_00848	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	198	65
e-chain	118	JGI_0000001-H6	Aigarchaeota	2265089299	J0000001H6DRAFT_00045	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	258	85
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264942887	A255F10DRAFT_01358	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	402	133
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264868182	A471O8DRAFT_01183	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	462	153

e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870890	A471C3DRAFT_00392	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	462	153
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089620	A471E16DRAFT_00130	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	543	180
e-chain	259	JGI_0000090-C20	Marinimicrobia	2265005630	J0000090C20DRAFT_00075	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1002	333
e-chain	161	JGI_0000039-O11	Marinimicrobia	2265097083	J0000039O11DRAFT_00037	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1002	333
e-chain	169	_JGI_0000059-L03	Marinimicrobia	2265099521	J0000059L03DRAFT_00424	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1002	333
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265112757	A252E7DRAFT_00234	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1005	334
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077558	AB164A22DRAFT_00400	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1008	335
e-chain	139	SCGC_AAA257-C11	BRC1	2264938318	A257C11DRAFT_01822	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1014	337
e-chain	130	SCGC_AAA255-G05	Atribacteria	2265070909	A255G5DRAFT_00231	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1020	339
e-chain	231	JGI_0000079-L04	Atribacteria	2265078614	J0000079L04DRAFT_00172	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1020	339
e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264985238	A252I11DRAFT_01689	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1029	342
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265092909	J0000106J15DRAFT_00190	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1038	345
e-chain	113	JGI_0000001-B8	Aigarchaeota	2265088894	J0000001B8DRAFT_00090	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1119	372
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089987	A471E16DRAFT_00497	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1119	372
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090177	A471F17DRAFT_00114	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1119	372
e-chain	171	JGI_0000059-M03	Caldiserica	2264941413	J0000059M03DRAFT_00446	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1203	400
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506118608	A7023_0009.00002020	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1230	409
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264868158	A471O8DRAFT_01159	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1860	619
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868304	A471B5DRAFT_00018	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1860	619
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870509	A471C3DRAFT_00011	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1860	619
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120093	A471L13DRAFT_00179	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1860	619
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265122001	A471L14DRAFT_01078	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1860	619
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944793	A252L23DRAFT_00998	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	897	298
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093221	J0000106J15DRAFT_00502	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	903	300

e-chain	14	SCGC_AAA471-G05	Aigarchaeota	2265091099	A471G5DRAFT_00174	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	915	304
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264868029	A471O8DRAFT_01030	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	933	310
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264869232	A471B5DRAFT_00946	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	933	310
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869807	A471B23DRAFT_00564	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	933	310
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120893	A471L13DRAFT_00979	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	933	310
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121478	A471L14DRAFT_00555	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	933	310
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867296	A471O8DRAFT_00297	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	945	314
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868825	A471B5DRAFT_00539	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	945	314
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264870213	A471B23DRAFT_00970	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	945	314
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264871208	A471C3DRAFT_00710	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	945	314
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265119960	A471L13DRAFT_00046	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	945	314
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121389	A471L14DRAFT_00466	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	945	314
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264869102	A471B5DRAFT_00816	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	951	316
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121601	A471L14DRAFT_00678	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	951	316
e-chain	10	SCGC_AAA471-E14	Aigarchaeota	2265087890	A471E14DRAFT_00139	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	954	317
e-chain	112	JGI_0000001-A7	Aigarchaeota	2265088364	J0000001A07DRAFT_00051	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	954	317
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101616	A03L08DRAFT_00667	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	957	318
e-chain	62	SCGC_AAA252-K16	Cloacimonetes	2264987791	A252K16DRAFT_00684	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	960	319
e-chain	268	SCGC_AAA003-E22	Marinimicrobia	2265100453	A03E22DRAFT_00292	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	960	319
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264959064	A471M6DRAFT_01001	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	972	323
e-chain	292	JGI_0000106-N7	Calescamantes	2264964753	J00000106N7DRAFT_00353	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	972	323
e-chain	294	JGI_0000106-P5	Calescamantes	2264966170	J0000106P5DRAFT_00409	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	972	323
e-chain	777	SCGC_AAA160-I06	Marinimicrobia	2265102562	DRAFT_00356	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	972	323
e-chain	778	SCGC_AAA298-D23	Marinimicrobia	2265103288	DRAFT_00074	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	972	323

e-chain	780	SCGC_AAA160-B08	Marinimicrobia	2265105293	DRAFT_00141	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	972	323
e-chain	221	SCGC_AB-179-E04	Thaumarchaeota	2265107848	AB179E04DRAFT_00242	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	972	323
e-chain	119	SCGC_AAA255-A06	Aminicenantes	2265032455	A255A6DRAFT_01049	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	975	324
e-chain	125	SCGC_AAA255-E05	Aminicenantes	2265032772	A255E5DRAFT_00029	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	975	324
e-chain	33	SCGC_AAA252-B17	Aminicenantes	2265040038	A252B17DRAFT_00738	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	975	324
e-chain	38	SCGC_AAA252-D06	Aminicenantes	2265041816	A252D6DRAFT_00547	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	975	324
e-chain	59	SCGC_AAA252-J21	Aminicenantes	2265055859	A252J21DRAFT_01177	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	975	324
e-chain	67	SCGC_AAA252-L10	Aminicenantes	2265061058	A252L10DRAFT_00686	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	975	324
e-chain	779	SCGC_AAA160-C11	Marinimicrobia	2265104779	DRAFT_00512	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	975	324
e-chain	181	SCGC_AAA011-A05	Marinimicrobia	2265005359	A11A5DRAFT_00154	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	981	326
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093070	J0000106J15DRAFT_00351	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	981	326
e-chain	164	JGI_0000059-D20	Marinimicrobia	2265098226	J0000059D20DRAFT_00431	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	981	326
e-chain	165	JGI_0000059-E23	Marinimicrobia	2265098685	J0000059E23DRAFT_00172	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	981	326
e-chain	81	SCGC_AAA252-O03	Aminicenantes	2265063149	A252O3DRAFT_00307	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1191	396
e-chain	170	JGI_0000059-L07	Cloacimonetes	2265116586	J0000059L07DRAFT_00174	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1191	396
e-chain	125	SCGC_AAA255-E05	Aminicenantes	2265033566	A255E5DRAFT_00824	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	31	SCGC_AAA252-A02	Aminicenantes	2265038449	A252A2DRAFT_00263	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	33	SCGC_AAA252-B17	Aminicenantes	2265039854	A252B17DRAFT_00553	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	38	SCGC_AAA252-D06	Aminicenantes	2265042503	A252D6DRAFT_01235	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	46	SCGC_AAA252-F08	Aminicenantes	2265044969	A252F8DRAFT_00481	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	48	SCGC_AAA252-G05	Aminicenantes	2265047726	A252G5DRAFT_01081	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	49	SCGC_AAA252-G06	Aminicenantes	2265049030	A252G6DRAFT_01266	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	54	SCGC_AAA252-I14	Aminicenantes	2265051713	A252I14DRAFT_01219	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	58	SCGC_AAA252-J15	Aminicenantes	2265054511	A252J15DRAFT_00777	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397

e-chain	61	SCGC_AAA252-K07	Aminicenantes	2265057895	A252K7DRAFT_00732	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	67	SCGC_AAA252-L10	Aminicenantes	2265061460	A252L10DRAFT_01088	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	83	SCGC_AAA252-O09	Aminicenantes	2265065471	A252O9DRAFT_01586	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	84	SCGC_AAA252-O10	Aminicenantes	2265066730	A252O10DRAFT_01072	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	93	SCGC_AAA252-P19	Aminicenantes	2265070067	A252P19DRAFT_00847	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265114017	A252E7DRAFT_01494	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1194	397
e-chain	78	SCGC_AAA252-N14	Cloacimonetes	2518602817	A252N14PERMD_00725	electron transfer flavoprotein alpha subunit apoprotein	flavoprotein-quinone oxidoreductase	1182	393
e-chain	37	SCGC_AAA252-D03	Cloacimonetes	2264977324	A252D3DRAFT_00853	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264985124	A252I11DRAFT_01575	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	56	SCGC_AAA252-I16	Cloacimonetes	2264986089	A252I16DRAFT_00718	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	68	SCGC_AAA252-L13	Cloacimonetes	2264989633	A252L13DRAFT_00723	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	72	SCGC_AAA252-M08	Cloacimonetes	2264991039	A252M8DRAFT_01311	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	74	SCGC_AAA252-N05	Cloacimonetes	2264991528	A252N5DRAFT_00245	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	77	SCGC_AAA252-N11	Cloacimonetes	2264992341	A252N11DRAFT_00176	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	79	SCGC_AAA252-N17	Cloacimonetes	2264997038	A252N17DRAFT_01752	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	87	SCGC_AAA252-O17	Cloacimonetes	2264999894	A252O17DRAFT_01528	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	90	SCGC_AAA252-P10	Cloacimonetes	2265002706	A252P10DRAFT_01737	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	92	SCGC_AAA252-P14	Cloacimonetes	2265003579	A252P14DRAFT_00647	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1182	393
e-chain	145	SCGC_AAA257-K07	Latescibacteria	2265109268	A257K7DRAFT_00546	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1200	399
e-chain	71	SCGC_AAA252-M02	Atribacteria	2265080351	A252M2DRAFT_00200	Aspartate oxidase	succinate/fumarate-quinone oxidoreductase	597	198
e-chain	291	JGI_0000106-N5	Calecamantes	2264964392	J0000106N5DRAFT_00600	Aspartate oxidase	succinate/fumarate-quinone oxidoreductase	855	284
e-chain	164	JGI_0000059-D20	Marinimicrobia	2265098238	J0000059D20DRAFT_00443	Aspartate oxidase	succinate/fumarate-quinone oxidoreductase	876	291
e-chain	149	JGI_0000039-D08	Marinimicrobia	2265096310	J0000039D08DRAFT_00620	Aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1194	397
e-chain	232	JGI_0000079-F20	Atribacteria	2265080054	J0000079F20DRAFT_00588	Aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1566	521

e-chain	181	SCGC_AAA011-A05	Marinimicrobia	2265005374	A11A5DRAFT_00169	Aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1569	522
e-chain	146	SCGC_AAA257-N23	Marinimicrobia	2265094998	A257N23DRAFT_00218	Aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1569	522
e-chain	165	JGI_0000059-E23	Marinimicrobia	2265098695	J0000059E23DRAFT_00182	Aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1569	522
e-chain	40	SCGC_AAA252-D18	Aminicenantes	2265044448	A252D18DRAFT_01376	Aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	130	SCGC_AAA255-G05	Atribacteria	2265071779	A255G5DRAFT_01110	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1269	422
e-chain	206	SCGC_AAA011-E11	Diapherotrites	2264949360	A11E11DRAFT_00668	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1443	480
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101006	A03L08DRAFT_00057	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1569	522
e-chain	119	SCGC_AAA255-A06	Aminicenantes	2265032738	A255A6DRAFT_01333	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	36	SCGC_AAA252-C08	Aminicenantes	2265041227	A252C8DRAFT_00994	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	47	SCGC_AAA252-G04	Aminicenantes	2265046626	A252G4DRAFT_00988	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	49	SCGC_AAA252-G06	Aminicenantes	2265048879	A252G6DRAFT_01115	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	54	SCGC_AAA252-I14	Aminicenantes	2265050703	A252I14DRAFT_00207	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	57	SCGC_AAA252-J09	Aminicenantes	2265053708	A252J9DRAFT_01802	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	63	SCGC_AAA252-K18	Aminicenantes	2265059385	A252K18DRAFT_00511	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	83	SCGC_AAA252-O09	Aminicenantes	2265065628	A252O9DRAFT_01743	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	85	SCGC_AAA252-O11	Aminicenantes	2265067723	A252O11DRAFT_00904	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	93	SCGC_AAA252-P19	Aminicenantes	2265070472	A252P19DRAFT_01252	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1575	524
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077504	AB164A22DRAFT_00346	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1587	528
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958693	A471M6DRAFT_00630	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1590	529
e-chain	286	JGI_0000106-I17	Calescamantes	2264960970	J0000106I17DRAFT_00108	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1590	529
e-chain	289	JGI_0000106-J16	Calescamantes	2264962395	J0000106J16DRAFT_00578	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1590	529
e-chain	290	JGI_0000106-M22	Calescamantes	2264962923	J0000106M22DRAFT_00096	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1590	529
e-chain	292	JGI_0000106-N7	Calescamantes	2264964427	J00000106N7DRAFT_00027	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1590	529
e-chain	294	JGI_0000106-P5	Calescamantes	2264966047	J0000106P5DRAFT_00286	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1590	529

e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265110714	A252B13DRAFT_00429	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1602	533
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265114020	A252E7DRAFT_01497	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1602	533
e-chain	155	JGI_0000039-I17	Cloacimonetes	2265115302	J0000039I17DRAFT_00214	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1602	533
e-chain	166	JGI_0000059-F17	Cloacimonetes	2265116294	J0000059F17DRAFT_00195	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1602	533
e-chain	170	JGI_0000059-L07	Cloacimonetes	2265116932	J0000059L07DRAFT_00522	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1602	533
e-chain	173	JGI_0000059-P03	Cloacimonetes	2265117430	J0000059P03DRAFT_00396	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1602	533
e-chain	73	SCGC_AAA252-M09	BRC1	2264940539	A252M9DRAFT_00556	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1605	534
e-chain	115	JGI_0000001-D5	Fervidibacteria	2264954705	J0000001D05DRAFT_00519	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1608	535
e-chain	117	JGI_0000001-G10	Fervidibacteria	2264955151	J0000001G10DRAFT_00404	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1623	540
e-chain	145	SCGC_AAA257-K07	Latescibacteria	2265109263	A257K7DRAFT_00541	L-aspartate oxidase (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1680	559
e-chain	147	SCGC_AAA257-O07	Omnitrophica	2265027250	A257O7DRAFT_00470	succinate dehydrogenase or fumarate reductase, flavoprotein subunit, Bacillus subtilis subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1740	579
e-chain	139	SCGC_AAA257-C11	BRC1	2264937091	A257C11DRAFT_00593	succinate dehydrogenase or fumarate reductase, flavoprotein subunit, Bacillus subtilis subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1875	624
e-chain	778	SCGC_AAA298-D23	Marinimicrobia	2265103413	DRAFT_00199	succinate dehydrogenase or fumarate reductase, flavoprotein subunit, Bacillus subtilis subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1923	640
e-chain	780	SCGC_AAA160-B08	Marinimicrobia	2265105163	DRAFT_00011	succinate dehydrogenase or fumarate reductase, flavoprotein subunit, Bacillus subtilis subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1923	640
e-chain	779	SCGC_AAA160-C11	Marinimicrobia	2265105084	DRAFT_00817	succinate dehydrogenase or fumarate reductase, flavoprotein subunit, Bacillus subtilis subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1956	651
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867695	A471O8DRAFT_00696	succinate dehydrogenase or fumarate reductase, flavoprotein subunitGram-negative/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1701	566
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868333	A471B5DRAFT_00047	succinate dehydrogenase or fumarate reductase, flavoprotein subunitGram-negative/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1701	566
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869446	A471B23DRAFT_00203	succinate dehydrogenase or fumarate reductase, flavoprotein subunitGram-negative/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1701	566
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120118	A471L13DRAFT_00204	succinate dehydrogenase or fumarate reductase, flavoprotein subunitGram-negative/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1701	566

e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121278	A471L14DRAFT_00355	succinate dehydrogenase or fumarate reductase, flavoprotein subunitGram-negative/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1701	566
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089613	A471E16DRAFT_00123	succinate dehydrogenase or fumarate reductase, flavoprotein subunitGram-negative/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1716	571
e-chain	1	SCGC_AAA471-A16	Aigarchaeota	2265092582	A471A16DRAFT_00142	succinate dehydrogenase or fumarate reductase, flavoprotein subunitGram-negative/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1716	571
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090461	A471F17DRAFT_00398	succinate dehydrogenase or fumarate reductase, flavoprotein subunitGram-negative/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1731	576
e-chain	4	SCGC_AAA471-B22	Aigarchaeota	2518715882	A471B22DRAFT_00220	succinate dehydrogenase or fumarate reductase, flavoprotein subunitGram-negative/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1731	576
e-chain	114	JGI_0000001-B9	Fervidibacteria	2264954055	J0000001B09DRAFT_00730	succinate dehydrogenase, flavoprotein subunit, E. coli/mitochondrial subgroup (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1779	592
e-chain	222	JGI_0000014-C22	Aerophobetes	2264943322	J0000014C22DRAFT_00430	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	succinate/fumarate-quinone oxidoreductase	1320	439
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264959659	A471M6DRAFT_01596	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	succinate/fumarate-quinone oxidoreductase	873	290
e-chain	73	SCGC_AAA252-M09	BRC1	2264940685	A252M9DRAFT_00702	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	succinate/fumarate-quinone oxidoreductase	939	312
e-chain	282	JGI_0000106_G12	Calescamantes	2264960140	J0000106G12DRAFT_00325	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1305	434
e-chain	292	JGI_0000106-N7	Calescamantes	2264965637	J00000106N7DRAFT_01237	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1395	464
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117866	A7O23_0002.00000270	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1713	570
e-chain	286	JGI_0000106-I17	Calescamantes	2264961265	J0000106I17DRAFT_00404	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1731	576
e-chain	289	JGI_0000106-J16	Calescamantes	2264962732	J0000106J16DRAFT_00915	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1731	576
e-chain	292	JGI_0000106-N7	Calescamantes	2264964772	J00000106N7DRAFT_00372	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1731	576
e-chain	294	JGI_0000106-P5	Calescamantes	2264965930	J0000106P5DRAFT_00169	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1731	576
e-chain	130	SCGC_AAA255-G05	Atribacteria	2265071328	A255G5DRAFT_00652	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1764	587

e-chain	216	SCGC_AB-164-D21	Atribacteria	2265075696	AB164D21DR AFT_00364	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1782	593
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017722	J0000077K19 DRAFT_00733	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1800	599
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265100990	A03L08DRAFT _00041	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1830	609
e-chain	223	JGI_J00000014-A15	Aerophobetes	2264943567	J0000014A15 DRAFT_00192	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.3.99.1)	succinate/fumarate-quinone oxidoreductase	1842	613
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264946098	A252L23DRAF T_02303	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit (EC:1.4.3.16)	succinate/fumarate-quinone oxidoreductase	1311	436
e-chain	78	SCGC_AAA252-N14	Cloacimonetes	2518602797	A252N14PER MD_00705	electron transfer flavoprotein beta subunit	flavoprotein-quinone oxidoreductase	693	230
e-chain	147	SCGC_AAA257-O07	Omnitrophica	2265027446	A257O7DRAFT _00666	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	1032	343
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870891	A471C3DRAFT _00393	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	441	146
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089574	A471E16DRAF T_00084	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	507	168
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073568	A255N14DRAF T_01168	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	699	232
e-chain	113	JGI_0000001-B8	Aigarchaeota	2265088895	J0000001B8D RAFT_00091	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	1203	400
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090178	A471F17DRAF T_00115	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	1227	408
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073577	A255N14DRAF T_01177	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	252	83
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073229	A255N14DRAF T_00827	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	258	85
e-chain	66	SCGC_AAA252-L05	Cloacimonetes	2264988583	A252L5DRAFT _00380	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	267	89
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264870893	A471C3DRAFT _00395	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	318	106
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073230	A255N14DRAF T_00828	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	336	111
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093796	J0000106J15D RAFT_01077	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	522	173
e-chain	50	SCGC_AAA252-G07	Cloacimonetes	2264981858	A252G7DRAFT _00485	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	591	197
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264869233	A471B5DRAFT _00947	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	600	199
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120894	A471L13DRAF T_00980	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	624	207
e-chain	171	JGI_0000059-M03	Caldiserica	2264941412	J0000059M03 DRAFT_00445	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	645	214
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867297	A471O8DRAFT _00298	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	741	246

e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868824	A471B5DRAFT_00538	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	741	246
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264870214	A471B23DRAFT_00971	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	741	246
e-chain	6	SCGC_AAA471-C03	Geoarchaeota	2264871209	A471C3DRAFT_00711	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	741	246
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265119959	A471L13DRAFT_00045	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	741	246
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121390	A471L14DRAFT_00467	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	741	246
e-chain	181	SCGC_AAA011-A05	Marinimicrobia	2265005360	A11A5DRAFT_00155	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	747	248
e-chain	164	JGI_0000059-D20	Marinimicrobia	2265098227	J0000059D20DRAFT_00432	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	747	248
e-chain	165	JGI_0000059-E23	Marinimicrobia	2265098686	J0000059E23DRAFT_00173	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	747	248
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264959063	A471M6DRAFT_01000	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	759	252
e-chain	292	JGI_0000106-N7	Calescamantes	2264964754	J00000106N7DRAFT_00354	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	759	252
e-chain	294	JGI_0000106-P5	Calescamantes	2264966169	J0000106P5DRAFT_00408	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	759	252
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093220	J0000106J15DRAFT_00501	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	762	253
e-chain	778	SCGC_AAA298-D23	Marinimicrobia	2265103289	DRAFT_00075	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	765	254
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101617	A03L08DRAFT_00668	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	768	255
e-chain	777	SCGC_AAA160-I06	Marinimicrobia	2265102563	DRAFT_00357	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	768	255
e-chain	779	SCGC_AAA160-C11	Marinimicrobia	2265104780	DRAFT_00513	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	768	255
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264869101	A471B5DRAFT_00815	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	771	256
e-chain	780	SCGC_AAA160-B08	Marinimicrobia	2265105292	DRAFT_00140	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	771	256
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121600	A471L14DRAFT_00677	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	771	256
e-chain	11	SCGC_AAA471-E16	Aigarchaeota	2265089519	A471E16DRAFT_00029	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	777	258
e-chain	14	SCGC_AAA471-G05	Aigarchaeota	2265091513	A471G5DRAFT_00588	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	777	258
e-chain	123	SCGC_AAA255-C20	Cloacimonetes	2264972597	A255C20DRAFT_00292	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	43	SCGC_AAA252-E13	Cloacimonetes	2264977934	A252E13DRAFT_00471	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	51	SCGC_AAA252-G08	Cloacimonetes	2264982429	A252G8DRAFT_00558	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259

e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264984268	A252I11DRAFT_00716	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	56	SCGC_AAA252-I16	Cloacimonetes	2264986400	A252I16DRAFT_01030	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	66	SCGC_AAA252-L05	Cloacimonetes	2264988571	A252L5DRAFT_00368	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	72	SCGC_AAA252-M08	Cloacimonetes	2264990147	A252M8DRAFT_00417	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	79	SCGC_AAA252-N17	Cloacimonetes	2264997044	A252N17DRAFT_01758	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	80	SCGC_AAA252-O02	Cloacimonetes	2264998201	A252O2DRAFT_00689	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	87	SCGC_AAA252-O17	Cloacimonetes	2264999776	A252O17DRAFT_01410	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	268	SCGC_AAA003-E22	Marinimicrobia	2265100454	A03E22DRAFT_00293	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	780	259
e-chain	119	SCGC_AAA255-A06	Aminicenantes	2265032456	A255A6DRAFT_01050	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	783	260
e-chain	125	SCGC_AAA255-E05	Aminicenantes	2265032773	A255E5DRAFT_00030	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	783	260
e-chain	33	SCGC_AAA252-B17	Aminicenantes	2265040039	A252B17DRAFT_00739	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	783	260
e-chain	38	SCGC_AAA252-D06	Aminicenantes	2265041817	A252D6DRAFT_00548	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	783	260
e-chain	59	SCGC_AAA252-J21	Aminicenantes	2265055858	A252J21DRAFT_01176	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	783	260
e-chain	67	SCGC_AAA252-L10	Aminicenantes	2265061059	A252L10DRAFT_00687	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	783	260
e-chain	221	SCGC_AB-179-E04	Thaumarchaeota	2265107847	AB179E04DRAFT_00241	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	783	260
e-chain	125	SCGC_AAA255-E05	Aminicenantes	2265033567	A255E5DRAFT_00825	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	31	SCGC_AAA252-A02	Aminicenantes	2265038448	A252A2DRAFT_00262	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	33	SCGC_AAA252-B17	Aminicenantes	2265039855	A252B17DRAFT_00554	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	38	SCGC_AAA252-D06	Aminicenantes	2265042502	A252D6DRAFT_01234	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	46	SCGC_AAA252-F08	Aminicenantes	2265044968	A252F8DRAFT_00480	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	49	SCGC_AAA252-G06	Aminicenantes	2265049031	A252G6DRAFT_01267	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	54	SCGC_AAA252-I14	Aminicenantes	2265050543	A252I14DRAFT_00047	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	58	SCGC_AAA252-J15	Aminicenantes	2265054510	A252J15DRAFT_00776	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	61	SCGC_AAA252-K07	Aminicenantes	2265057894	A252K7DRAFT_00731	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261

e-chain	67	SCGC_AAA252-L10	Aminicenantes	2265061459	A252L10DRAFT_01087	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	81	SCGC_AAA252-O03	Aminicenantes	2265063148	A252O3DRAFT_00306	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	83	SCGC_AAA252-O09	Aminicenantes	2265065470	A252O9DRAFT_01585	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	84	SCGC_AAA252-O10	Aminicenantes	2265066729	A252O10DRAFT_01071	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	93	SCGC_AAA252-P19	Aminicenantes	2265070066	A252P19DRAFT_00846	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	231	JGI_0000079-L04	Atribacteria	2265079017	J0000079L04DRAFT_00575	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	10	SCGC_AAA471-E14	Aigarchaeota	2265087889	A471E14DRAFT_00138	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	112	JGI_0000001-A7	Aigarchaeota	2265088365	J0000001A07DRAFT_00052	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	159	JGI_0000039-M09	Cloacimonetes	2265115420	J0000039M09DRAFT_00034	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	173	JGI_0000059-P03	Cloacimonetes	2265117108	J0000059P03DRAFT_00074	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	174	JGI_0000059-P11	Cloacimonetes	2265117806	J0000059P11DRAFT_00318	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	786	261
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264868028	A471O8DRAFT_01029	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	789	262
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869806	A471B23DRAFT_00563	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	789	262
e-chain	130	SCGC_AAA255-G05	Atribacteria	2265070910	A255G5DRAFT_00232	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	789	262
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073569	A255N14DRAFT_01169	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	789	262
e-chain	231	JGI_0000079-L04	Atribacteria	2265078613	J0000079L04DRAFT_00171	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	789	262
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093071	J0000106J15DRAFT_00352	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	789	262
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121479	A471L14DRAFT_00556	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	789	262
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264941732	A255F10DRAFT_00198	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	792	263
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944813	A252L23DRAFT_01018	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	792	263
e-chain	145	SCGC_AAA257-K07	Latescibacteria	2265109267	A257K7DRAFT_00545	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	792	263
e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265110712	A252B13DRAFT_00427	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	792	263
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265114018	A252E7DRAFT_01495	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	792	263
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265112756	A252E7DRAFT_00233	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	795	264

e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264985237	A252I11DRAFT_01688	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	798	265
e-chain	259	JGI_0000090-C20	Marinimicrobia	2265005629	J0000090C20DRAFT_00074	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	798	265
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077559	AB164A22DR AFT_00401	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	798	265
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265092910	J0000106J15DRAFT_00191	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	798	265
e-chain	161	JGI_0000039-O11	Marinimicrobia	2265097082	J0000039O11DRAFT_00036	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	798	265
e-chain	169	_JGI_0000059-L03	Marinimicrobia	2265099520	J0000059L03DRAFT_00423	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	798	265
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506118609	A7O23_0009.00002030	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	882	293
e-chain	139	SCGC_AAA257-C11	BRC1	2264938317	A257C11DRAFT_01821	Electron transfer flavoprotein, beta subunit	flavoprotein-quinone oxidoreductase	906	301
e-chain	162	JGI_0000039-P09	Cloacimonetes	2265115868	J0000039P09DRAFT_00027	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	837	278
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264941733	A255F10DRAFT_00199	Electron transfer flavoprotein, alpha subunit	flavoprotein-quinone oxidoreductase	834	278
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944812	A252L23DRAFT_01017	Predicted NADH:ubiquinone oxidoreductase, subunit RnfB	flavoprotein-quinone oxidoreductase	408	135
e-chain	281	JGI_0000106-F11	Aenigmarchaea	2264946357	J0000106F11DRAFT_00143	Coenzyme F420-reducing hydrogenase, gamma subunit	NADH-quinone oxidoreductase	483	160
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073795	A255N14DRAFT_01396	Coenzyme F420-reducing hydrogenase, gamma subunit	NADH-quinone oxidoreductase	756	251
e-chain	215	SCGC_AB-164-P05	Atribacteria	2265075003	AB164P05DRAFT_00183	Coenzyme F420-reducing hydrogenase, gamma subunit	NADH-quinone oxidoreductase	756	251
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030352	A11A17DRAFT_01202	Coenzyme F420-reducing hydrogenase, gamma subunit	NADH-quinone oxidoreductase	768	255
e-chain	10	SCGC_AAA471-E14	Aigarchaeota	2265088249	A471E14DRAFT_00498	Coenzyme F420-reducing hydrogenase, gamma subunit	NADH-quinone oxidoreductase	777	258
e-chain	60	SCGC_AAA252-K06	Aminicenantes	2265056351	A252K6DRAFT_00370	Coenzyme F420-reducing hydrogenase, gamma subunit	NADH-quinone oxidoreductase	780	259
e-chain	273	JGI_0000014_D06	Cloacimonetes	2265118390	J0000014D06DRAFT_00538	Coenzyme F420-reducing hydrogenase, gamma subunit	NADH-quinone oxidoreductase	783	260
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265113557	A252E7DRAFT_01034	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.1.2)	NADH-quinone oxidoreductase	546	181
e-chain	172	JGI_0000059-N10	Hydrogenedentes	2265009558	J0000059N10DRAFT_00054	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.1.2)	NADH-quinone oxidoreductase	549	182
e-chain	37	SCGC_AAA252-D03	Cloacimonetes	2264977299	A252D3DRAFT_00828	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337
e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264984834	A252I11DRAFT_01284	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337
e-chain	56	SCGC_AAA252-I16	Cloacimonetes	2264986323	A252I16DRAFT_00953	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337
e-chain	72	SCGC_AAA252-M08	Cloacimonetes	2264989737	A252M8DRAFT_00006	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337

e-chain	74	SCGC_AAA252-N05	Cloacimonetes	2264991986	A252N5DRAFT_00703	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337
e-chain	79	SCGC_AAA252-N17	Cloacimonetes	2264997097	A252N17DRAFT_01811	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337
e-chain	87	SCGC_AAA252-O17	Cloacimonetes	2264998513	A252O17DRAFT_00146	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337
e-chain	90	SCGC_AAA252-P10	Cloacimonetes	2265001470	A252P10DRAFT_00499	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337
e-chain	92	SCGC_AAA252-P14	Cloacimonetes	2265003654	A252P14DRAFT_00722	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337
e-chain	259	JGI_0000090-C20	Marinimicrobia	2265006048	J0000090C20DRAFT_00494	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	750	250
e-chain	161	JGI_0000039-O11	Marinimicrobia	2265097215	J0000039O11DRAFT_00169	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	972	323
e-chain	234	JGI_0000077-B04	Marinimicrobia	2265099777	J0000077B04DRAFT_00095	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	972	323
e-chain	236	JGI_J00000077-D07	Hydrogenedentes	2265011013	J0000077D07DRAFT_00464	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	978	325
e-chain	126	SCGC_AAA255-E10	Aminicenantes	2265034000	A255E10DRAFT_00412	Coenzyme F420-reducing hydrogenase, gamma subunit (EC:1.12.99.-)	NADH-quinone oxidoreductase	981	326
e-chain	78	SCGC_AAA252-N14	Cloacimonetes	2518603297	A252N14PERMD_01205	F420-non-reducing hydrogenase subunit G (EC:1.12.99.-)	NADH-quinone oxidoreductase	1014	337
e-chain	139	SCGC_AAA257-C11	BRC1	2264938204	A257C11DRAFT_01707	hydrogenase (NiFe) small subunit (hydA) (EC:1.12.7.2)	NADH-quinone oxidoreductase	900	299
e-chain	78	SCGC_AAA252-N14	Cloacimonetes	2518602965	A252N14PERMD_00873	Membrane bound hydrogenase subunit mbhJ	NADH-quinone oxidoreductase	420	139
e-chain	4	SCGC_AAA471-B22	Aigarchaeota	2518715895	A471B22DRAFT_00233	NADH dehydrogenase subunit B (EC 1.6.5.3) (EC:1.6.5.3)	NADH-quinone oxidoreductase	657	218
e-chain	130	SCGC_AAA255-G05	Atribacteria	2265071502	A255G5DRAFT_00832	NADH-quinone oxidoreductase, B subunit	NADH-quinone oxidoreductase	450	149
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073776	A255N14DRAFT_01377	NADH-quinone oxidoreductase, B subunit	NADH-quinone oxidoreductase	450	149
e-chain	218	SCGC_AB-164-M20	Atribacteria	2265076727	AB164M20DRAFT_00074	NADH-quinone oxidoreductase, B subunit	NADH-quinone oxidoreductase	450	149
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077617	AB164A22DRAFT_00459	NADH-quinone oxidoreductase, B subunit	NADH-quinone oxidoreductase	450	149
e-chain	231	JGI_0000079-L04	Atribacteria	2265079434	J0000079L04DRAFT_00997	NADH-quinone oxidoreductase, B subunit	NADH-quinone oxidoreductase	462	153
e-chain	114	JGI_0000001-B9	Fervidibacteria	2264953357	J0000001B09DRAFT_00030	NADH-quinone oxidoreductase, B subunit	NADH-quinone oxidoreductase	660	219
e-chain	117	JGI_0000001-G10	Fervidibacteria	2264954756	J0000001G10DRAFT_00008	NADH-quinone oxidoreductase, B subunit	NADH-quinone oxidoreductase	660	219
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030255	A11A17DRAFT_01105	NADH-quinone oxidoreductase, B subunit	NADH-quinone oxidoreductase	777	258
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030532	A11A17DRAFT_01382	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	480	159
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101136	A03L08DRAFT_00187	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	510	169

e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867697	A471O8DRAFT_00698	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	549	182
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868335	A471B5DRAFT_00049	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	549	182
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869448	A471B23DRAFT_00205	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	549	182
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120120	A471L13DRAFT_00206	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	549	182
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121276	A471L14DRAFT_00353	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	549	182
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264959466	A471M6DRAFT_01403	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	561	186
e-chain	282	JGI_0000106_G12	Calescamantes	2264960105	J0000106G12DRAFT_00290	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	561	186
e-chain	289	JGI_0000106-J16	Calescamantes	2264961937	J0000106J16DRAFT_00120	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	561	186
e-chain	290	JGI_0000106-M22	Calescamantes	2264963161	J0000106M22DRAFT_00334	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	561	186
e-chain	292	JGI_0000106-N7	Calescamantes	2264964505	J00000106N7DRAFT_00105	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	561	186
e-chain	10	SCGC_AAA471-E14	Aigarchaeota	2265088252	A471E14DRAFT_00501	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	645	214
e-chain	118	JGI_0000001-H6	Aigarchaeota	2265089393	J0000001H6DRAFT_00139	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	657	218
e-chain	14	SCGC_AAA471-G05	Aigarchaeota	2265091466	A471G5DRAFT_00541	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	657	218
e-chain	18	SCGC_AAA471-J07	Aigarchaeota	2265091876	A471J7DRAFT_00016	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	657	218
e-chain	288	JGI_0000106-J15	Aigarchaeota	2265093479	J0000106J15DRAFT_00760	NADH-quinone oxidoreductase, B subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	660	219
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077813	AB164A22DRAFT_00659	NADH:ubiquinone oxidoreductase 20 kD subunit and related Fe-S oxidoreductases (EC:1.6.5.3)	NADH-quinone oxidoreductase	426	141
e-chain	145	SCGC_AAA257-K07	Latescibacteria	2265109688	A257K7DRAFT_00966	NADH:ubiquinone oxidoreductase 20 kD subunit and related Fe-S oxidoreductases (EC:1.6.5.3)	NADH-quinone oxidoreductase	558	185
e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265111379	A252B13DRAFT_01095	NADH:ubiquinone oxidoreductase 20 kD subunit and related Fe-S oxidoreductases (EC:1.6.5.3)	NADH-quinone oxidoreductase	558	185
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117345	A7023_0001.00001060	NADH:ubiquinone oxidoreductase 20 kD subunit and related Fe-S oxidoreductases (EC:1.6.5.3)	NADH-quinone oxidoreductase	558	185
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944681	A252L23DRAFT_00886	NADH:ubiquinone oxidoreductase 20 kD subunit and related Fe-S oxidoreductases (EC:1.6.5.3)	NADH-quinone oxidoreductase	561	186
e-chain	7	SCGC_AAA471-D06	Fervidibacteria	2264966458	A471D6DRAFT_00197	NADH:ubiquinone oxidoreductase 20 kD subunit and related Fe-S oxidoreductases (EC:1.6.5.3)	NADH-quinone oxidoreductase	618	205
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944154	A252L23DRAFT_00358	Ni,Fe-hydrogenase I small subunit	NADH-quinone oxidoreductase	1086	361

e-chain	121	SCGC_AAA255-B16	Cloacimonetes	2264971419	A255B16DRAFT_00128	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	37	SCGC_AAA252-D03	Cloacimonetes	2264977333	A252D3DRAFT_00862	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	43	SCGC_AAA252-E13	Cloacimonetes	2264978236	A252E13DRAFT_00775	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	51	SCGC_AAA252-G08	Cloacimonetes	2264982408	A252G8DRAFT_00537	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264983960	A252I11DRAFT_00408	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	74	SCGC_AAA252-N05	Cloacimonetes	2264991331	A252N5DRAFT_00048	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	79	SCGC_AAA252-N17	Cloacimonetes	2264996519	A252N17DRAFT_01232	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	87	SCGC_AAA252-O17	Cloacimonetes	2264998594	A252O17DRAFT_00227	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	259	JGI_0000090-C20	Marinimicrobia	2265005966	J0000090C20DRAFT_00412	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	146	SCGC_AAA257-N23	Marinimicrobia	2265095624	A257N23DRAFT_00844	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	161	JGI_0000039-O11	Marinimicrobia	2265097551	J0000039O11DRAFT_00506	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	420	139
e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265110577	A252B13DRAFT_00292	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	435	144
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265113284	A252E7DRAFT_00761	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	435	144
e-chain	233	JGI_0000079-D21	Synergistetes	2265107492	J0000079D21DRAFT_00910	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	462	153
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017048	J0000077K19DRAFT_00056	Ni,Fe-hydrogenase III small subunit	NADH-quinone oxidoreductase	474	157
e-chain	115	JGI_0000001-D5	Fervidibacteria	2264954478	J0000001D05DRAFT_00292	Ni,Fe-hydrogenase III small subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	531	176
e-chain	129	SCGC_AAA255-F10	Aerophobetes	2264942040	A255F10DRAFT_00508	Ni,Fe-hydrogenase III small subunit (EC:1.6.99.3)	NADH-quinone oxidoreductase	414	137
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944358	A252L23DRAFT_00563	Ni,Fe-hydrogenase III small subunit (EC:1.6.99.3)	NADH-quinone oxidoreductase	420	139
e-chain	145	SCGC_AAA257-K07	Latescibacteria	2265108995	A257K7DRAFT_00273	Ni,Fe-hydrogenase III small subunit (EC:1.6.99.3)	NADH-quinone oxidoreductase	432	143
e-chain	77	SCGC_AAA252-N11	Cloacimonetes	2264992324	A252N11DRAFT_00159	Ni,Fe-hydrogenase III small subunit (EC:1.6.99.3)	NADH-quinone oxidoreductase	453	150
e-chain	139	SCGC_AAA257-C11	BRC1	2264937090	A257C11DRAFT_00592	succinate dehydrogenase (or fumarate reductase) cytochrome b subunit, b558 family	succinate/fumarate-quinone oxidoreductase	630	209
e-chain	147	SCGC_AAA257-O07	Omnitrophica	2265027251	A257O7DRAFT_00471	succinate dehydrogenase (or fumarate reductase) cytochrome b subunit, b558 family	succinate/fumarate-quinone oxidoreductase	636	211
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117868	A7O23_0002.00000290	Succinate dehydrogenase, hydrophobic anchor subunit	succinate/fumarate-quinone oxidoreductase	357	118
e-chain	113	JGI_0000001-B8	Aigarchaeota	2265088855	J0000001B8DRAFT_00051	Succinate dehydrogenase, hydrophobic anchor subunit	succinate/fumarate-quinone oxidoreductase	372	123

e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090308	A471F17DRAFT_00245	Succinate dehydrogenase, hydrophobic anchor subunit	succinate/fumarate-quinone oxidoreductase	372	123
e-chain	1	SCGC_AAA471-A16	Aigarchaeota	2265092481	A471A16DRAFT_00041	Succinate dehydrogenase, hydrophobic anchor subunit	succinate/fumarate-quinone oxidoreductase	372	123
e-chain	783	SCGC_AAA007-O23	Thaumarchaeota	2506117867	A7O23_0002.00000280	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit	succinate/fumarate-quinone oxidoreductase	429	142
e-chain	113	JGI_0000001-B8	Aigarchaeota	2265088856	J0000001B8DRAFT_00052	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit	succinate/fumarate-quinone oxidoreductase	441	146
e-chain	13	SCGC_AAA471-F17	Aigarchaeota	2265090307	A471F17DRAFT_00244	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit	succinate/fumarate-quinone oxidoreductase	441	146
e-chain	1	SCGC_AAA471-A16	Aigarchaeota	2265092480	A471A16DRAFT_00040	Succinate dehydrogenase/fumarate reductase, cytochrome b subunit	succinate/fumarate-quinone oxidoreductase	441	146
e-chain	70	SCGC_AAA252-L23	Aerophobetes	2264944606	A252L23DRAFT_00811	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3138	1045
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030535	A11A17DRAFT_01385	NADH-quinone oxidoreductase, F subunit	NADH-quinone oxidoreductase	1815	604
e-chain	164	JGI_0000059-D20	Marinimicrobia	2265097862	J0000059D20DRAFT_00066	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.12.1.2)	NADH-quinone oxidoreductase	1704	567
e-chain	130	SCGC_AAA255-G05	Atribacteria	2265072142	A255G5DRAFT_01476	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1746	581
e-chain	219	SCGC_AB-164-A22	Atribacteria	2265077613	AB164A22DRAFT_00455	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1746	581
e-chain	122	SCGC_AAA255-C06	Acetothermia	2265016718	A255C6DRAFT_00501	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1755	584
e-chain	60	SCGC_AAA252-K06	Aminicenantes	2265056357	A252K6DRAFT_00376	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1833	610
e-chain	152	JGI_0000039-G13	Cloacimonetes	2265115034	J0000039G13DRAFT_00187	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1860	619
e-chain	173	JGI_0000059-P03	Cloacimonetes	2265117366	J0000059P03DRAFT_00332	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1860	619
e-chain	174	JGI_0000059-P11	Cloacimonetes	2265117632	J0000059P11DRAFT_00144	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1860	619
e-chain	274	JGI_0000014-D18	Cloacimonetes	2265118734	J0000014D18DRAFT_00341	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1896	631
e-chain	145	SCGC_AAA257-K07	Latescibacteria	2265109032	A257K7DRAFT_00310	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	2361	786
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017600	J0000077K19DRAFT_00610	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	2460	819
e-chain	222	JGI_0000014-C22	Aerophobetes	2264943085	J0000014C22DRAFT_00193	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1854	617
e-chain	172	JGI_0000059-N10	Hydrogenedentes	2265009556	J0000059N10DRAFT_00052	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.12.1.2)	NADH-quinone oxidoreductase	1773	590
e-chain	293	JGI_0000106-O11	Thermotogae	2265108202	J0000106O11DRAFT_00262	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	2007	668
e-chain	128	SCGC_AAA255-F09	Aminicenantes	2265035287	A255F9DRAFT_00257	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	2904	967
e-chain	63	SCGC_AAA252-K18	Aminicenantes	2265059213	A252K18DRAFT_00339	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3153	1050

e-chain	81	SCGC_AAA252-O03	Aminicenantes	2265063501	A252O3DRAFT_00660	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3153	1050
e-chain	91	SCGC_AAA252-P13	Aminicenantes	2265068765	A252P13DRAFT_00584	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3153	1050
e-chain	274	JGI_0000014-D18	Cloacimonetes	2265118744	J0000014D18DRAFT_00351	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	3156	1051
e-chain	240	JGI_0000077-K19	Armatimonadetes	2265017694	J0000077K19DRAFT_00704	NADH-quinone oxidoreductase, F subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1296	431
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264959475	A471M6DRAFT_01412	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1293	430
e-chain	283	JGI_0000106_H18	Calescamantes	2264960597	J0000106H18DRAFT_00269	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1293	430
e-chain	282	JGI_0000106_G12	Calescamantes	2264960113	J0000106G12DRAFT_00298	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1296	431
e-chain	289	JGI_0000106-J16	Calescamantes	2264961945	J0000106J16DRAFT_00128	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1296	431
e-chain	290	JGI_0000106-M22	Calescamantes	2264963153	J0000106M22DRAFT_00326	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1296	431
e-chain	292	JGI_0000106-N7	Calescamantes	2264964513	J00000106N7DRAFT_00113	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1296	431
e-chain	121	SCGC_AAA255-B16	Cloacimonetes	2264972281	A255B16DRAFT_00991	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1428	476
e-chain	276	JGI_0000014-K11	Cloacimonetes	2265119623	J0000014K11DRAFT_00310	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1452	484
e-chain	152	JGI_0000039-G13	Cloacimonetes	2265115032	J0000039G13DRAFT_00185	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1563	520
e-chain	166	JGI_0000059-F17	Cloacimonetes	2265116376	J0000059F17DRAFT_00278	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1563	520
e-chain	173	JGI_0000059-P03	Cloacimonetes	2265117364	J0000059P03DRAFT_00330	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1563	520
e-chain	174	JGI_0000059-P11	Cloacimonetes	2265117634	J0000059P11DRAFT_00146	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1563	520
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073513	A255N14DRAFT_01112	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1587	528
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030189	A11A17DRAFT_01039	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1617	538
e-chain	52	SCGC_AAA252-G21	Aminicenantes	2265050236	A252G21DRAFT_01094	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1644	547
e-chain	123	SCGC_AAA255-C20	Cloacimonetes	2264972623	A255C20DRAFT_00318	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	1002	333
e-chain	166	JGI_0000059-F17	Cloacimonetes	2265116374	J0000059F17DRAFT_00276	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	1284	427
e-chain	43	SCGC_AAA252-E13	Cloacimonetes	2264978519	A252E13DRAFT_01058	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	44	SCGC_AAA252-E20	Cloacimonetes	2264979877	A252E20DRAFT_00804	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	45	SCGC_AAA252-F02	Cloacimonetes	2264980637	A252F2DRAFT_00040	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592

e-chain	53	SCGC_AAA252-I11	Cloacimonetes	2264983924	A252I11DRAFT_00372	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	56	SCGC_AAA252-I16	Cloacimonetes	2264986742	A252I16DRAFT_01372	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	62	SCGC_AAA252-K16	Cloacimonetes	2264987776	A252K16DRAFT_00669	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	77	SCGC_AAA252-N11	Cloacimonetes	2264992180	A252N11DRAFT_00015	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	79	SCGC_AAA252-N17	Cloacimonetes	2264996428	A252N17DRAFT_01141	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	80	SCGC_AAA252-O02	Cloacimonetes	2264997678	A252O2DRAFT_00166	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	87	SCGC_AAA252-O17	Cloacimonetes	2264999934	A252O17DRAFT_01568	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	90	SCGC_AAA252-P10	Cloacimonetes	2265002549	A252P10DRAFT_01579	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	126	SCGC_AAA255-E10	Aminicenantes	2265033845	A255E10DRAFT_00257	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1551	516
e-chain	233	JGI_0000079-D21	Synergistetes	2265106946	J0000079D21DRAFT_00361	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1755	584
e-chain	149	JGI_0000039-D08	Marinimicrobia	2265095743	J0000039D08DRAFT_00051	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1947	648
e-chain	172	JGI_0000059-N10	Hydrogenedentes	2265009595	J0000059N10DRAFT_00091	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	2181	726
e-chain	34	SCGC_AAA252-B19	Omnitrophica	2265031126	A252B19DRAFT_00178	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	1209	402
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265113559	A252E7DRAFT_01036	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit	NADH-quinone oxidoreductase	1443	480
e-chain	178	SCGC_AAA036-E14	Parcubacteria	2265013804	A36E14DRAFT_00491	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1104	367
e-chain	280	JGI_0000106-B10	EM3	2264966619	J0000106B10DRAFT_00125	NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit (EC:1.6.5.3)	NADH-quinone oxidoreductase	1158	385
e-chain	136	SCGC_AAA255-N14	Atribacteria	2265073731	A255N14DRAFT_01332	Cytochrome bd-type quinol oxidase, subunit 1	cytochrome/quinol oxidase-aerobic	1371	456
e-chain	139	SCGC_AAA257-C11	BRC1	2264936998	A257C11DRAFT_00500	Cytochrome bd-type quinol oxidase, subunit 1 (EC:1.10.3.-)	cytochrome/quinol oxidase-aerobic	1341	446
e-chain	73	SCGC_AAA252-M09	BRC1	2264940537	A252M9DRAFT_00554	Cytochrome bd-type quinol oxidase, subunit 1 (EC:1.10.3.-)	cytochrome/quinol oxidase-aerobic	1440	479
e-chain	115	JGI_0000001-D5	Fervidibacteria	2264954492	J0000001D05DRAFT_00306	Cytochrome bd-type quinol oxidase, subunit 1	cytochrome/quinol oxidase-aerobic	1314	437
e-chain	172	JGI_0000059-N10	Hydrogenedentes	2265009594	J0000059N10DRAFT_00090	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	1779	592
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265030188	A11A17DRAFT_01038	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	1746	581
e-chain	152	JGI_0000039-G13	Cloacimonetes	2265115031	J0000039G13DRAFT_00184	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	1785	594
e-chain	173	JGI_0000059-P03	Cloacimonetes	2265117363	J0000059P03DRAFT_00329	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	1785	594

e-chain	174	JGI_0000059-P11	Cloacimonetes	2265117635	J0000059P11 DRAFT_00147	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	1785	594
e-chain	166	JGI_0000059-F17	Cloacimonetes	2265116377	J0000059F17D RAFT_00279	Iron only hydrogenase large subunit, C-terminal domain (EC:1.6.5.3)	NADH-quinone oxidoreductase	1785	594
e-chain	34	SCGC_AAA252-B19	Omnitrophica	2265031127	A252B19DRAFT_00179	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	1899	632
e-chain	55	SCGC_AAA252-I15	Euryarchaeota	2265081183	A252I15DRAFT_00128	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	1725	574
e-chain	178	SCGC_AAA036-E14	Parcubacteria	2265013803	A36E14DRAFT_00490	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	1722	573
e-chain	164	JGI_0000059-D20	Marinimicrobia	2265097861	J0000059D20 DRAFT_00065	hydrogenase, Fe-only (EC:1.6.5.3)	NADH-quinone oxidoreductase	1932	643
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867693	A471O8DRAFT_00694	Fumarate reductase subunit C	succinate/fumarate-quinone oxidoreductase	360	119
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868331	A471B5DRAFT_00045	Fumarate reductase subunit C	succinate/fumarate-quinone oxidoreductase	360	119
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869444	A471B23DRAFT_00201	Fumarate reductase subunit C	succinate/fumarate-quinone oxidoreductase	360	119
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120116	A471L13DRAFT_00202	Fumarate reductase subunit C	succinate/fumarate-quinone oxidoreductase	360	119
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121280	A471L14DRAFT_00357	Fumarate reductase subunit C	succinate/fumarate-quinone oxidoreductase	360	119
e-chain	29	SCGC_AAA471-O08	Geoarchaeota	2264867692	A471O8DRAFT_00693	Fumarate reductase subunit D.	succinate/fumarate-quinone oxidoreductase	336	111
e-chain	2	SCGC_AAA471-B05	Geoarchaeota	2264868330	A471B5DRAFT_00044	Fumarate reductase subunit D.	succinate/fumarate-quinone oxidoreductase	336	111
e-chain	5	SCGC_AAA471-B23	Geoarchaeota	2264869443	A471B23DRAFT_00200	Fumarate reductase subunit D.	succinate/fumarate-quinone oxidoreductase	336	111
e-chain	22	SCGC_AAA471-L13	Geoarchaeota	2265120115	A471L13DRAFT_00201	Fumarate reductase subunit D.	succinate/fumarate-quinone oxidoreductase	336	111
e-chain	23	SCGC_AAA471-L14	Geoarchaeota	2265121281	A471L14DRAFT_00358	Fumarate reductase subunit D.	succinate/fumarate-quinone oxidoreductase	336	111
e-chain	73	SCGC_AAA252-M09	BRC1	2264940536	A252M9DRAFT_00553	cytochrome d oxidase, subunit II (cydB) (EC:1.10.3.-)	cytochrome/quinol oxidase-aerobic	1014	337
e-chain	39	SCGC_AAA252-D10	Latescibacteria	2265112297	A252D10DRAFT_00435	cytochrome d oxidase, subunit II (cydB) (EC:1.10.3.-)	cytochrome/quinol oxidase-aerobic	1014	337
e-chain	139	SCGC_AAA257-C11	BRC1	2264936997	A257C11DRAFT_00499	cytochrome d oxidase, subunit II (cydB) (EC:1.10.3.-)	cytochrome/quinol oxidase-aerobic	1017	338
e-chain	150	JGI_0000039-E15	Marinimicrobia	2265096714	J0000039E15 DRAFT_00129	Aspartate oxidase	succinate/fumarate-quinone oxidoreductase	144	47
e-chain	263	SCGC_AAA011-A17	Omnitrophica	2265029374	A11A17DRAFT_00224	caa(3)-type oxidase, subunit IV	cytochrome/quinol oxidase-aerobic	300	99
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101175	A03L08DRAFT_00226	caa(3)-type oxidase, subunit IV	cytochrome/quinol oxidase-aerobic	357	118
e-chain	780	SCGC_AAA160-B08	Marinimicrobia	2265105680	DRAFT_00529	caa(3)-type oxidase, subunit IV	cytochrome/quinol oxidase-aerobic	417	138
e-chain	781	SCGC_AAA076-M08	Marinimicrobia	2265106550	DRAFT_00404	caa(3)-type oxidase, subunit IV (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	417	138

e-chain	778	SCGC_AAA298-D23	Marinimicrobia	2265103955	DRAFT_00743	caa(3)-type oxidase, subunit IV (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	420	139
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958087	A471M6DRAFT_00024	hypothetical protein	cytochrome/quinol oxidase-aerobic	312	103
e-chain	283	JGI_0000106_H18	Calescamantes	2264960648	J0000106H18 DRAFT_00320	hypothetical protein	cytochrome/quinol oxidase-aerobic	312	103
e-chain	289	JGI_0000106-J16	Calescamantes	2264962709	J0000106J16DRAFT_00892	hypothetical protein	cytochrome/quinol oxidase-aerobic	312	103
e-chain	290	JGI_0000106-M22	Calescamantes	2264963735	J0000106M22 DRAFT_00908	hypothetical protein	cytochrome/quinol oxidase-aerobic	312	103
e-chain	291	JGI_0000106-N5	Calescamantes	2264963968	J0000106N5DRAFT_00176	hypothetical protein	cytochrome/quinol oxidase-aerobic	312	103
e-chain	292	JGI_0000106-N7	Calescamantes	2264964907	J00000106N7 DRAFT_00507	hypothetical protein	cytochrome/quinol oxidase-aerobic	312	103
e-chain	14	SCGC_AAA471-G05	Aigarchaeota	2265091265	A471G5DRAFT_00340	hypothetical protein (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	285	94
e-chain	777	SCGC_AAA160-I06	Marinimicrobia	2265103149	DRAFT_00945	Predicted small integral membrane protein (EC:1.9.3.1)	cytochrome/quinol oxidase-aerobic	417	138
e-chain	282	JGI_0000106_G12	Calescamantes	2264960315	J0000106G12 DRAFT_00501	Polysulphide reductase	quinol-cytochrome oxidoreductase	1578	525
e-chain	292	JGI_0000106-N7	Calescamantes	2264965140	J00000106N7 DRAFT_00740	Polysulphide reductase	quinol-cytochrome oxidoreductase	1578	525
e-chain	184	SCGC_AAA011-L16	Omnitrophica	2265028431	A11L16DRAFT_00054	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1098	365
e-chain	171	JGI_0000059-M03	Caldiserica	2264941000	J0000059M03 DRAFT_00032	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1158	385
e-chain	778	SCGC_AAA298-D23	Marinimicrobia	2265103946	DRAFT_00734	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1380	459
e-chain	777	SCGC_AAA160-I06	Marinimicrobia	2265103158	DRAFT_00954	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1389	462
e-chain	780	SCGC_AAA160-B08	Marinimicrobia	2265105689	DRAFT_00538	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1389	462
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264959629	A471M6DRAFT_01566	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1410	469
e-chain	282	JGI_0000106_G12	Calescamantes	2264960195	J0000106G12 DRAFT_00381	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1410	469
e-chain	289	JGI_0000106-J16	Calescamantes	2264962782	J0000106J16DRAFT_00965	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1410	469
e-chain	290	JGI_0000106-M22	Calescamantes	2264963716	J0000106M22 DRAFT_00889	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1410	469
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958981	A471M6DRAFT_00918	Polysulphide reductase (EC:1.2.7.-)	quinol-cytochrome oxidoreductase	1590	529
e-chain	268	SCGC_AAA003-E22	Marinimicrobia	2265100452	A03E22DRAFT_00291	Dehydrogenases (flavoproteins) (EC:1.5.5.1)	flavoprotein-quinone oxidoreductase	1674	557
e-chain	156	JGI_0000039-J10	Hydrogenedentes	2265009266	J0000039J10DRAFT_00106	Cbb3-type cytochrome oxidase component FixQ.	cytochrome/quinol oxidase-aerobic	177	58
e-chain	237	JGI_J00000077-O16	Hydrogenedentes	2265011859	J0000077O16 DRAFT_00093	Cbb3-type cytochrome oxidase component FixQ.	cytochrome/quinol oxidase-aerobic	177	58

e-chain	234	JGI_0000077-B04	Marinimicrobia	2265100158	J0000077B04 DRAFT_00477	Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrA	NADH-quinone oxidoreductase	384	128
e-chain	777	SCGC_AAA160-I06	Marinimicrobia	2265102283	DRAFT_00076	Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrA	NADH-quinone oxidoreductase	738	245
e-chain	269	SCGC_AAA003-L8	Marinimicrobia	2265101561	A03L08DRAFT _00612	NADH:ubiquinone oxidoreductase, Na(+)- translocating, A subunit (EC:1.6.5.-)	NADH-quinone oxidoreductase	1386	461
e-chain	777	SCGC_AAA160-I06	Marinimicrobia	2265103172	DRAFT_00968	Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrA	NADH-quinone oxidoreductase	699	232
e-chain	778	SCGC_AAA298-D23	Marinimicrobia	2265104099	DRAFT_00888	NADH:ubiquinone oxidoreductase, Na(+)- translocating, A subunit (EC:1.6.5.-)	NADH-quinone oxidoreductase	1332	443
e-chain	779	SCGC_AAA160-C11	Marinimicrobia	2265104454	DRAFT_00186	NADH:ubiquinone oxidoreductase, Na(+)- translocating, A subunit (EC:1.6.5.-)	NADH-quinone oxidoreductase	1338	445
e-chain	780	SCGC_AAA160-B08	Marinimicrobia	2265105934	DRAFT_00783	NADH:ubiquinone oxidoreductase, Na(+)- translocating, A subunit (EC:1.6.5.-)	NADH-quinone oxidoreductase	1338	445
e-chain	149	JGI_0000039-D08	Marinimicrobia	2265096584	J0000039D08 DRAFT_00894	NADH:ubiquinone oxidoreductase, Na(+)- translocating, A subunit (EC:1.6.5.-)	NADH-quinone oxidoreductase	1353	450
e-chain	169	_JGI_0000059-L03	Marinimicrobia	2265099202	J0000059L03D RAFT_00105	NADH:ubiquinone oxidoreductase, Na(+)- translocating, A subunit (EC:1.6.5.-)	NADH-quinone oxidoreductase	1353	450
e-chain	234	JGI_0000077-B04	Marinimicrobia	2265099949	J0000077B04 DRAFT_00267	NADH:ubiquinone oxidoreductase, Na(+)- translocating, A subunit (EC:1.6.5.-)	NADH-quinone oxidoreductase	987	328
e-chain	32	SCGC_AAA252-B13	Latescibacteria	2265110753	A252B13DRAF T_00468	Predicted NADH:ubiquinone oxidoreductase, subunit RnfC	NADH-quinone oxidoreductase	1368	455
e-chain	25	SCGC_AAA471-M6	Calescamantes	2264958355	A471M6DRAF T_00292	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G)	NADH-quinone oxidoreductase	1281	426
e-chain	283	JGI_0000106_H18	Calescamantes	2264960847	J0000106H18 DRAFT_00519	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G)	NADH-quinone oxidoreductase	1281	426
e-chain	289	JGI_0000106-J16	Calescamantes	2264962084	J0000106J16D RAFT_00267	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G)	NADH-quinone oxidoreductase	1281	426
e-chain	290	JGI_0000106-M22	Calescamantes	2264963044	J0000106M22 DRAFT_00217	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G)	NADH-quinone oxidoreductase	1281	426
e-chain	42	SCGC_AAA252-E07	Latescibacteria	2265113558	A252E7DRAFT _01035	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G) (EC:1.12.1.2)	NADH-quinone oxidoreductase	705	234
e-chain	172	JGI_0000059-N10	Hydrogenedentes	2265009557	J0000059N10 DRAFT_00053	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G) (EC:1.12.1.2)	NADH-quinone oxidoreductase	717	238
e-chain	232	JGI_0000079-F20	Atribacteria	2265079584	J0000079F20D RAFT_00117	NADH dehydrogenase/NADH:ubiquinone oxidoreductase 75 kD subunit (chain G)	NADH-quinone oxidoreductase	573	190
e-chain	231	JGI_0000079-L04	Atribacteria	2265079458	J0000079L04D RAFT_01021	NADH:ubiquinone oxidoreductase, NADH- binding (51 kD) subunit	NADH-quinone oxidoreductase	447	148

Supplementary Table 7 | Archaeal PurP genes in Microgenomates (OP11) SAGs.

IMG annotation	SAG	aa	gene ID	locustag
ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases) (EC:6.3.4.-)	Microgenomates SCGC AAA011-A19 (193)	337	2265021031	A11A19DRAFT_00045
ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases) (EC:6.3.4.-)	Microgenomates SCGC AAA011-L05 (205)	337	2265023504	A11L05DRAFT_00194
ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases) (EC:6.3.4.-)	Microgenomates SCGC AAA011-L6 (266)	338	2265025581	A11L06DRAFT_00258

Supplementary Table 8 | Bacterial sigma factors in archaeal SAGs.

locustag	Gene Product Name	aa	Genome	COG	Pfam	KO
A11E11DRAFT_00143	RNA polymerase sigma factor, sigma-70 family	391	Diapherotrites SCGC AAA011-E11 (206)	COG0568	pfam04542 <<>>pfam04545	
A11N19DRAFT_00074	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	231	Diapherotrites SCGC AAA011-N19 (272)	COG1595	pfam04542	
A11G17DRAFT_01061	RNA polymerase sigma factor, sigma-70 family	179	Nanoarchaeota SCGC AAA011-G17 (198)	COG1595	pfam04542 <<>>pfam08281	KO:K03088
A11K22DRAFT_00364	RNA polymerase sigma factor, sigma-70 family	181	Nanoarchaeota SCGC AAA011-K22 (202)	COG1595	pfam04542 <<>>pfam08281	KO:K03088
A11L22DRAFT_00377	DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog	179	Nanoarchaeota SCGC AAA011-L22 (192)	COG1595	pfam04542 <<>>pfam08281	KO:K03088
A252I15DRAFT_00551	RNA polymerase sigma factor, sigma-70 family	218	Euryarchaeota SCGC AAA252-I15 (55)	COG1595	pfam08281	

Supplementary Table 9 | Genes encoding the stringent response (ppGpp) in archaeal SAGs. Full length homologs of RelA/SpoT were found in Diapherotrites (pMC2A384) and Nanoarchaeota. The Diapherotrites genome contains two RelA/SpoT homologs.

IMG annotation	SAG	aa	gene ID	locustag
Guanosine polyphosphate pyrophosphohydrolases/synthetases (COG0317); TIGR00691 (p)ppGpp synthetase, RelA/SpoT family	Diapherotrites SCGC AAA011-E11 (206)	618	2264948981	A11E11DRA FT_00289
Guanosine polyphosphate pyrophosphohydrolases/synthetases (COG0317); TIGR00691 (p)ppGpp synthetase, RelA/SpoT family	Diapherotrites SCGC AAA011-E11 (206) [2nd copy]	596	2264949432	A11E11DRA FT_00740
TIGR00691 (p)ppGpp synthetase, RelA/SpoT family; COG317	Nanoarchaeota SCGC AAA011-G17 (198)	629	2264950849	A11G17DRA FT_00028

Supplementary Table 10 | Phylogenetic anchoring details. (a) Phylogenetically anchored reads are separated into the number of newly assigned reads versus number of re-assigned (r.a.) reads. The 15 most anchored phyla are shown against the top 15 taxa from which reads were re-assigned. (b) Approximately half of the reads in the 475 metagenomes could not be assigned (defined here as not being classified beyond the domain level) even after including our MDM reference genomes. Shown are the unassigned reads for the 19 metagenomes detailed in Fig. 5A, which mostly belong to the taxonomic level of cellular organisms and the domain Bacteria.

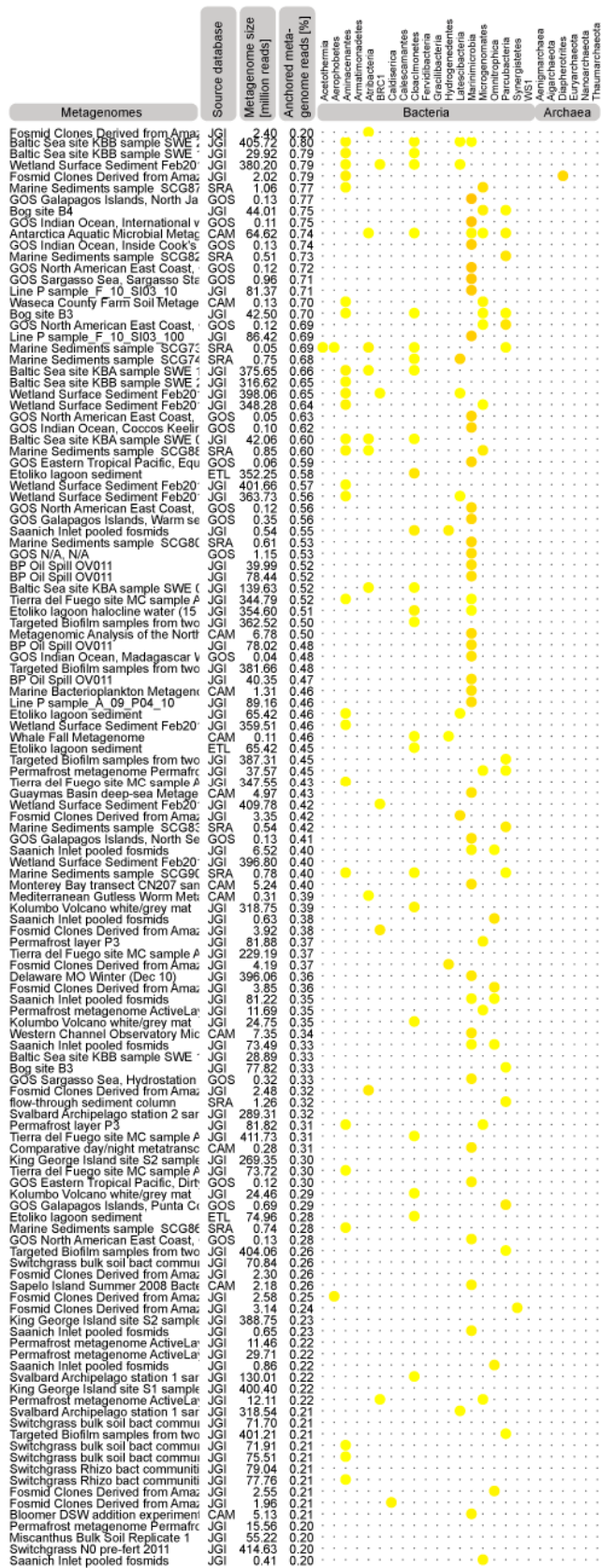
a

	Sum	Newly assigned	r.a. from Proteobacteria	r.a. from Bacteroidetes/Chlorobi group	r.a. from Firmicutes	r.a. from Opisthokonta	r.a. from Euryarchaeota	r.a. from Fibrobacteres/Acidobacteria group	r.a. from Planctomycetes	r.a. from Actinobacteria	r.a. from Chloroflexi	r.a. from Cyanobacteria	r.a. from Chlamydiae/Verrucomicrobia group	r.a. from Viridiplantae	r.a. from Spirochaetes	r.a. from Thaumarchaeota	r.a. from Crenarchaeota	r.a. from Others
Sum	100%	77.61	5.66	5.26	2.91	1.90	1.35	0.77	0.76	0.61	0.47	0.45	0.34	0.31	0.29	0.24	0.23	0.85
Marinimicrobia (SAR406)	39.96	30.96	2.28	3.36	0.86	0.65	0.20	0.42	0.33	0.16	0.08	0.17	0.10	0.10	0.03	0.02	0.15	
Aminacenantetes (OP8)	11.13	8.81	0.83	0.47	0.17	0.18	0.13	0.18	0.04	0.08	0.03	0.03	0.03	0.04	0.01	0.01	0.08	
Cloacimonetes (WWE1)	7.27	5.64	0.45	0.38	0.17	0.15	0.09	0.07	0.04	0.03	0.05	0.05	0.02	0.03	0.02	0.00	0.07	
Fervidibacteria (Oct-Spa1-106)	6.32	4.83	0.34	0.09	0.24	0.27	0.06	0.03	0.06	0.07	0.08	0.05	0.04	0.04	0.03	0.00	0.07	
Omnitrophica (OP3)	5.78	4.32	0.44	0.40	0.22	0.09	0.02	0.01	0.06	0.01	0.00	0.03	0.02	0.00	0.02	0.00	0.12	
Parcubacteria (OD1)	4.78	3.97	0.19	0.04	0.23	0.06	0.05	0.01	0.00	0.04	0.08	0.02	0.01	0.03	0.02	0.00	0.03	
Latescibacteria (WS3)	4.43	3.46	0.24	0.27	0.08	0.09	0.04	0.01	0.07	0.04	0.02	0.01	0.03	0.02	0.01	0.00	0.03	
Microgenomates (OP11)	4.18	3.38	0.24	0.03	0.16	0.07	0.06	0.00	0.02	0.04	0.06	0.04	0.01	0.01	0.01	0.00	0.02	
Atribacteria (OP9)	3.99	3.03	0.15	0.05	0.42	0.10	0.07	0.01	0.01	0.06	0.01	0.01	0.00	0.01	0.03	0.00	0.04	
Aerophobetes (CD12)	1.94	1.51	0.10	0.03	0.13	0.03	0.04	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.02	
candidate division BRC1	1.79	1.35	0.09	0.06	0.04	0.04	0.02	0.01	0.08	0.02	0.01	0.00	0.05	0.01	0.01	0.00	0.00	
Hydrogenedentes (NKB19)	1.30	1.01	0.11	0.03	0.03	0.03	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	
Thaumarchaeota	1.25	1.08	0.03	0.01	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.04	0.01	
Aigarchaeota (pSL4)	1.01	0.69	0.02	0.00	0.01	0.01	0.03	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.18	0.06	
Others	4.88	3.56	0.14	0.04	0.12	0.11	0.54	0.00	0.03	0.03	0.01	0.01	0.02	0.01	0.01	0.02	0.21	

b

Metagenome	Total percentage of unassigned reads	Root	Cellular organisms						Viruses	No hits
			Bacteria	Archaea	Eukaryota					
Lake Sakinaw	59.12	1.40	29.51	23.64	0.00	0.44	0.01	4.12		
TA mother reactor	34.88	1.17	15.14	14.81	0.44	0.87	0.05	2.40		
TA biofilm	46.07	1.39	19.94	19.36	1.02	0.91	0.14	3.31		
GBS CS 85C sediment	44.98	1.68	17.62	14.80	0.28	1.42	0.45	8.73		
GBS 77S water	51.52	2.58	21.56	17.11	0.20	1.95	0.27	7.85		
Saanich Inlet pooled fosmids	38.17	1.13	17.60	8.17	2.92	1.19	0.09	7.07		
GOS Mangrove on Isabella Island	42.64	1.29	16.42	17.78	0.95	1.09	0.08	5.03		
Yellowstone Bison Hot Spring	55.21	1.68	21.89	20.02	0.38	1.40	0.34	9.50		
GBS CS 77C sediment	68.83	2.68	32.13	24.52	0.36	1.67	0.29	7.18		
GOS Sargasso Stations 3	36.09	1.35	17.65	12.02	0.19	0.69	0.09	4.10		
TA Sludge	66.41	2.11	27.76	25.89	0.56	1.20	0.16	8.73		
Line P A09P20-1000	30.07	1.31	10.37	15.47	0.33	0.95	0.07	1.57		
Rifle 2008 BG 13ft	72.18	0.77	17.14	13.94	1.81	0.88	0.11	37.53		
Peru Margin	28.91	0.83	10.34	12.10	0.15	0.38	0.08	5.03		
Marine Sediments sample SCG71	65.37	3.52	28.59	19.75	0.30	2.36	0.66	10.19		
Etoliko lagoon sediment	29.87	1.26	8.96	17.30	0.13	1.48	0.41	0.33		
GBS 77S sediment	35.45	1.90	14.44	16.69	0.19	1.65	0.47	0.11		
Biofilm from redox zones LI09_3	76.53	0.54	6.13	6.54	0.05	1.09	0.17	62.01		
Line P J09P20-500	42.74	1.85	21.20	14.35	0.09	0.86	0.09	4.30		
Average:	48.69	1.60	18.65	16.54	0.54	1.18	0.21	9.95		

Sample	Source database	Metagenome size [million reads]	Anchored meta- genomes [%]	Phylogenetic anchoring																									
				0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%	16%	17%	18%	19%	20%	21%	22%	23%	24%	25%
Lake Sakinaw	JGI	386.58	19.56																										
TA mother reactor	CAM	0.17	14.39																										
TA biotini	TAB	189.34	10.82																										
GBS CS 85C sediment	JGI	78.59	9.27																										
GBS 77S water	JGI	0.92	8.83																										
Saanchin Inlet pooled fosmids	GOS	0.14	6.24																										
GOS Mangrove on Isabela Island	GOS	0.47	5.18																										
Yellowstone Bison Hot Spring	JGI	86.63	4.63																										
GBS CS 77C sediment	GOS	0.36	4.38																										
GOS Sargasso Stations 3	TAS	214.99	4.30																										
TA Sludge	JGI	0.72	3.84																										
Saanchin Inlet pooled fosmids	JGI	0.68	3.58																										
Saanchin Inlet pooled fosmids	JGI	83.90	3.47																										
Rifle 2008 BG 13t	JGI	70.26	3.11																										
Line P A09P20-1000	JGI	83.96	3.39																										
Line P J08P26-500	JGI	85.44	3.32																										
Peru Margin	CAM	0.60	3.28																										
Saanchin Inlet pooled fosmids	JGI	1.21	3.04																										
Rifle 2008 BG 13t	JGI	222.00	3.01																										
Line P A09P20-500	JGI	91.36	2.97																										
Line P A09P04-500	JGI	86.19	2.96																										
Saanchin Inlet pooled fosmids	JGI	37.77	2.96																										
Line P A09P04-1000	JGI	0.92	2.91																										
Marine Sediments sample SCG71	SRA	64.46	2.85																										
Rifle 2008 BG 13t	JGI	0.7																											



Supplementary Table 12 | Physico-chemical measurements of sampling sites.

site	category	depth	temp	pH	O2	total H ₂ S	CH ₄	conductivity (EC)
		<i>m</i>	<i>°C</i>		<i>μM</i>	<i>μM</i>	<i>mM</i>	<i>μS/cm</i>
GBS	hot vent	0.3	81	7.07	41.7	0.00000048	n.d.	8440
EPR	diffuse-flow vent	2503	30	5.8	50	1000	0.01	53000
ETL	sediment	27.5	12.6	7.3	0	1200	n.d.	7310
SAK	brackish water	120	9.3	6.8	0	5761.7	3*	13200
TA	bioreactor	0	48	7	0	n.d.	20	n.d.
HSM	freshwater	100	n.d.	8	n.d.	n.d.	n.d.	750
TG	seawater	800	5	n.d.	120	n.d.	n.d.	33000
GOM	seawater	1	14.9	~8	315	n.d.	n.d.	38500
HOT	seawater	25	26.6	n.d.	204	n.d	n.d.	53000

Supplementary Table 13 | Conserved single copy gene sets for genome completeness and genome size estimation. Marker gene sets consisting of 139 bacterial conserved single copy genes (CSCGs ;shown in the top orange table) and 162 archaeal CSCGs (shown in the green bottom table) that were found to occur only once in at least 90% of all finished genomes in IMG were used (see Supplementary Methods).

PFAM ID	HMM name	HMM cutoff	Description of bacterial genes
PF03485	Arg_tRNA_synt_N	32.55	Arginyl tRNA synthetase N terminal domain
PF03484	B5	26.55	tRNA synthetase B5 domain
PF01121	CoaE	84.65	Dephospho-CoA kinase
PF03772	Competence	82.7	Competence protein
PF03602	Cons_hypoth95	38.8	Conserved hypothetical protein 95
PF06418	CTP_synt_N	208	CTP synthase N-terminus
PF02224	Cytidylate_kin	79.35	Cytidylate kinase
PF00712	DNA_pol3_beta	48.7	DNA polymerase III beta subunit, N-terminal domain
PF02767	DNA_pol3_beta_2	49.65	DNA polymerase III beta subunit, central domain
PF02768	DNA_pol3_beta_3	44.5	DNA polymerase III beta subunit, C-terminal domain
PF00035	dsrm	20.2	Double-stranded RNA binding motif
PF00889	EF_TS	108.35	Elongation factor TS
PF01176	eIF-1a	39.25	Translation initiation factor 1A / IF-1
PF00113	Enolase_C	206.35	Enolase, C-terminal TIM barrel domain
PF03952	Enolase_N	92.45	Enolase, N-terminal domain
PF06574	FAD_syn	73.25	FAD synthetase
PF03147	FDX-ACB	39.75	Ferredoxin-fold anticodon binding domain
PF01687	Flavokinase	59.3	Riboflavin kinase
PF02938	GAD	40.05	GAD domain
PF02527	GidB	77.85	rRNA small subunit methyltransferase G
PF00958	GMP_synt_C	61.3	GMP synthase C terminal domain
PF01025	GrpE	72.25	GrpE
PF01018	GTP1_OBG	98.9	GTP1/OBG
PF11987	IF-2	61.7	Translation-initiation factor 2
PF04760	IF2_N	29.95	Translation initiation factor IF-2, N-terminal region
PF00707	IF3_C	51.55	Translation initiation factor IF-3, C-terminal domain
PF05198	IF3_N	47.95	Translation initiation factor IF-3, N-terminal domain
PF01715	IPPT	129.15	IPP transferase
PF06421	LepA_C	79.7	GTP-binding protein LepA C-terminus
PF01795	Methyltransf_5	174.05	MraW methylase family
PF02873	MurB_C	46.15	UDP-N-acetylenolpyruvoylglucosamine reductase, C-terminal domain
PF08529	NusA_N	61.55	NusA N-terminal domain
PF02410	Oligomerisation	44.9	Oligomerisation domain
PF01195	Pept_tRNA_hydro	99.15	Peptidyl-tRNA hydrolase
PF01252	Peptidase_A8	61.35	Signal peptidase (SPase) II
PF00162	PGK	236	Phosphoglycerate kinase
PF02912	Phe_tRNA-synt_N	33.55	Aminoacyl tRNA synthetase class II, N-terminal domain
PF03726	PNPase	25.6	Polyribonucleotide nucleotidyltransferase, RNA binding domain

PF01416	PseudoU_synth_1	42.45	tRNA pseudouridine synthase
PF02033	RBFA	44.9	Ribosome-binding factor A
PF00154	RecA	276.55	recA bacterial DNA recombination protein
PF02132	RecR	21.25	RecR protein
PF00825	Ribonuclease_P	40.55	Ribonuclease P
PF00687	Ribosomal_L1	75.1	Ribosomal protein L1p/L10e family
PF00466	Ribosomal_L10	37.3	Ribosomal protein L10
PF00298	Ribosomal_L11	42.65	Ribosomal protein L11, RNA binding domain
PF03946	Ribosomal_L11_N	45.9	Ribosomal protein L11, N-terminal domain
PF00542	Ribosomal_L12	44	Ribosomal protein L7/L12 C-terminal domain
PF00572	Ribosomal_L13	81.3	Ribosomal protein L13
PF00238	Ribosomal_L14	79.55	Ribosomal protein L14p/L23e
PF00252	Ribosomal_L16	77.45	Ribosomal protein L16p/L10e
PF01196	Ribosomal_L17	55.75	Ribosomal protein L17
PF00828	Ribosomal_L18e	37.55	Ribosomal protein L18e/L15
PF00861	Ribosomal_L18p	54.85	Ribosomal L18p/L5e family
PF01245	Ribosomal_L19	71.7	Ribosomal protein L19
PF00181	Ribosomal_L2	52.45	Ribosomal Proteins L2, RNA binding domain
PF03947	Ribosomal_L2_C	87.3	Ribosomal Proteins L2, C-terminal domain
PF00453	Ribosomal_L20	70.5	Ribosomal protein L20
PF00829	Ribosomal_L21p	53.45	Ribosomal prokaryotic L21 protein
PF00237	Ribosomal_L22	57.6	Ribosomal protein L22p/L17e
PF00276	Ribosomal_L23	37.45	Ribosomal protein L23
PF01016	Ribosomal_L27	55.4	Ribosomal L27 protein
PF00830	Ribosomal_L28	36.8	Ribosomal L28 family
PF00831	Ribosomal_L29	33.7	Ribosomal L29 protein
PF00297	Ribosomal_L3	66.35	Ribosomal protein L3
PF01783	Ribosomal_L32p	28.2	Ribosomal L32p protein family
PF01632	Ribosomal_L35p	29.5	Ribosomal protein L35
PF00573	Ribosomal_L4	99.95	Ribosomal protein L4/L1 family
PF00281	Ribosomal_L5	34.55	Ribosomal protein L5
PF00673	Ribosomal_L5_C	61.25	ribosomal L5P family C-terminus
PF00347	Ribosomal_L6	53.25	Ribosomal protein L6
PF03948	Ribosomal_L9_C	37.95	Ribosomal protein L9, C-terminal domain
PF01281	Ribosomal_L9_N	30.55	Ribosomal protein L9, N-terminal domain
PF00338	Ribosomal_S10	58.05	Ribosomal protein S10p/S20e
PF00411	Ribosomal_S11	73.45	Ribosomal protein S11
PF00164	Ribosomal_S12	84.3	Ribosomal protein S12
PF00416	Ribosomal_S13	56.2	Ribosomal protein S13/S18
PF00312	Ribosomal_S15	41.7	Ribosomal protein S15
PF00886	Ribosomal_S16	36.25	Ribosomal protein S16
PF00366	Ribosomal_S17	41.3	Ribosomal protein S17
PF01084	Ribosomal_S18	36.3	Ribosomal protein S18
PF00203	Ribosomal_S19	54.9	Ribosomal protein S19

PF00318	Ribosomal_S2	133.15	Ribosomal protein S2
PF01649	Ribosomal_S20p	35.95	Ribosomal protein S20
PF00189	Ribosomal_S3_C	49.5	Ribosomal protein S3, C-terminal domain
PF00163	Ribosomal_S4	32.9	Ribosomal protein S4/S9 N-terminal domain
PF00333	Ribosomal_S5	42.55	Ribosomal protein S5, N-terminal domain
PF03719	Ribosomal_S5_C	44.9	Ribosomal protein S5, C-terminal domain
PF01250	Ribosomal_S6	45.2	Ribosomal protein S6
PF00177	Ribosomal_S7	95.55	Ribosomal protein S7p/S5e
PF00410	Ribosomal_S8	72.35	Ribosomal protein S8
PF00380	Ribosomal_S9	70.4	Ribosomal protein S9/S16
PF01782	RimM	30.65	RimM N-terminal domain
PF01000	RNA_pol_A_bac	34.3	RNA polymerase Rpb3/RpoA insert domain
PF03118	RNA_pol_A_CTD	38.9	Bacterial RNA polymerase, alpha chain C terminal domain
PF01193	RNA_pol_L	34.35	RNA polymerase Rpb3/Rpb11 dimerisation domain
PF04997	RNA_pol_Rpb1_1	157.25	RNA polymerase Rpb1, domain 1
PF00623	RNA_pol_Rpb1_2	79.8	RNA polymerase Rpb1, domain 2
PF04983	RNA_pol_Rpb1_3	41.85	RNA polymerase Rpb1, domain 3
PF05000	RNA_pol_Rpb1_4	24.6	RNA polymerase Rpb1, domain 4
PF04998	RNA_pol_Rpb1_5	119.4	RNA polymerase Rpb1, domain 5
PF04563	RNA_pol_Rpb2_1	39.95	RNA polymerase beta subunit
PF04561	RNA_pol_Rpb2_2	37.15	RNA polymerase Rpb2, domain 2
PF04565	RNA_pol_Rpb2_3	45.6	RNA polymerase Rpb2, domain 3
PF10385	RNA_pol_Rpb2_45	36.6	RNA polymerase beta subunit external 1 domain
PF00562	RNA_pol_Rpb2_6	258.45	RNA polymerase Rpb2, domain 6
PF04560	RNA_pol_Rpb2_7	46.45	RNA polymerase Rpb2, domain 7
PF01765	RRF	101.65	Ribosome recycling factor
PF07499	RuvA_C	16.6	RuvA, C-terminal domain
PF01330	RuvA_N	26.85	RuvA N terminal domain
PF05491	RuvB_C	49.5	Holliday junction DNA helicase ruvB C-terminus
PF02773	S-AdoMet_synt_C	108.9	S-adenosylmethionine synthetase, C-terminal domain
PF02772	S-AdoMet_synt_M	77.05	S-adenosylmethionine synthetase, central domain
PF00584	SecE	27.5	SecE/Sec61-gamma subunits of protein translocation complex
PF03840	SecG	29.2	Preprotein translocase SecG subunit
PF00344	SecY	180.7	SecY translocase
PF02403	Seryl_tRNA_N	43.15	Seryl-tRNA synthetase N-terminal domain
PF01668	SmpB	43.4	SmpB protein
PF02978	SRP_SPB	52.8	Signal peptide binding domain
PF00763	THF_DHG_CYH	62.2	Tetrahydrofolate dehydrogenase/cyclohydrolase, catalytic domain
PF02882	THF_DHG_CYH_C	103.7	Tetrahydrofolate dehydrogenase/cyclohydrolase, NAD(P)-binding domain
PF00121	TIM	140.2	Triosephosphate isomerase
PF08275	Toprim_N	58.6	DNA primase catalytic core, N-terminal domain
PF03461	TRCF	40.05	TRCF domain
PF05698	Trigger_C	55.15	Bacterial trigger factor protein (TF) C-terminus
PF05697	Trigger_N	61.8	Bacterial trigger factor protein (TF)

PF01746	tRNA_m1G_MT	80.75	tRNA (Guanine-1)-methyltransferase
PF00750	tRNA-synt_1d	120.85	tRNA synthetases class I (R)
PF01409	tRNA-synt_2d	161.15	tRNA synthetases class II core domain (F)
PF01509	TruB_N	83.55	TruB family pseudouridylate synthase (N terminal domain)
PF00627	UBA	13.3	UBA/TS-N domain
PF02130	UPF0054	59.35	Uncharacterized protein family UPF0054
PF02367	UPF0079	46.1	Uncharacterised P-loop hydrolase UPF0079
PF03652	UPF0081	60.75	Uncharacterised protein family (UPF0081)
PF12344	UvrB	33.75	Ultra-violet resistance protein B
PF08459	UvrC_HhH_N	81.9	UvrC Helix-hairpin-helix N-terminal
PF10458	Val_tRNA-synt_C	26.5	Valyl tRNA synthetase tRNA binding arm
PF06071	YchF-GTPase_C	61.9	Protein of unknown function (DUF933)
PF06689	zf-C4_ClpX	31.7	ClpX C4-type zinc finger

PFAM ID	HMM name	HMM cutoff	Description of archaeal genes
PF00113	Enolase_C	150.75	Enolase, C-terminal TIM barrel domain
PF00121	TIM	39.45	Triosephosphate isomerase
PF00137	ATP-synt_C	36.2	ATP synthase subunit C
PF00162	PGK	181.15	Phosphoglycerate kinase
PF00164	Ribosomal_S12	60.25	Ribosomal protein S12
PF00177	Ribosomal_S7	65.4	Ribosomal protein S7p/S5e
PF00181	Ribosomal_L2	20.35	Ribosomal Proteins L2, RNA binding domain
PF00189	Ribosomal_S3_C	24.9	Ribosomal protein S3, C-terminal domain
PF00203	Ribosomal_S19	39.75	Ribosomal protein S19
PF00237	Ribosomal_L22	51	Ribosomal protein L22p/L17e
PF00238	Ribosomal_L14	55	Ribosomal protein L14p/L23e
PF00252	Ribosomal_L16	46.05	Ribosomal protein L16p/L10e
PF00253	Ribosomal_S14	23.25	Ribosomal protein S14p/S29e
PF00276	Ribosomal_L23	23.7	Ribosomal protein L23
PF00281	Ribosomal_L5	25.75	Ribosomal protein L5
PF00297	Ribosomal_L3	114	Ribosomal protein L3
PF00298	Ribosomal_L11	31.7	Ribosomal protein L11, RNA binding domain
PF00312	Ribosomal_S15	28.8	Ribosomal protein S15
PF00318	Ribosomal_S2	55.9	Ribosomal protein S2
PF00327	Ribosomal_L30	24.05	Ribosomal protein L30p/L7e
PF00333	Ribosomal_S5	37.9	Ribosomal protein S5, N-terminal domain
PF00334	NDK	88.05	Nucleoside diphosphate kinase
PF00344	SecY	145.45	SecY translocase
PF00347	Ribosomal_L6	33.25	Ribosomal protein L6
PF00366	Ribosomal_S17	37.65	Ribosomal protein S17
PF00368	HMG-CoA_red	235	Hydroxymethylglutaryl-coenzyme A reductase
PF00380	Ribosomal_S9	50.5	Ribosomal protein S9/S16
PF00410	Ribosomal_S8	37	Ribosomal protein S8

PF00411	Ribosomal_S11	64.6	Ribosomal protein S11
PF00416	Ribosomal_S13	47.5	Ribosomal protein S13/S18
PF00466	Ribosomal_L10	32.4	Ribosomal protein L10
PF00562	RNA_pol_Rpb2_6	204.75	RNA polymerase Rpb2, domain 6
PF00572	Ribosomal_L13	39.35	Ribosomal protein L13
PF00573	Ribosomal_L4	63.9	Ribosomal protein L4/L1 family
PF00623	RNA_pol_Rpb1_2	107.9	RNA polymerase Rpb1, domain 2
PF00670	AdoHcyase_NAD	81.05	S-adenosyl-L-homocysteine hydrolase, NAD binding domain
PF00673	Ribosomal_L5_C	37.65	ribosomal L5P family C-terminus
PF00679	EFG_C	34.25	Elongation factor G C-terminus
PF00687	Ribosomal_L1	67.15	Ribosomal protein L1p/L10e family
PF00709	Adenylsucc_synt	130.9	Adenylosuccinate synthetase
PF00736	EF1_GNE	29.55	EF-1 guanine nucleotide exchange domain
PF00749	tRNA-synt_1c	120.7	tRNA synthetases class I (E and Q), catalytic domain
PF00750	tRNA-synt_1d	78.1	tRNA synthetases class I (R)
PF00752	XPG_N	30.15	XPG N-terminal domain
PF00827	Ribosomal_L15e	128.6	Ribosomal L15
PF00831	Ribosomal_L29	25.65	Ribosomal L29 protein
PF00832	Ribosomal_L39	28.85	Ribosomal L39 protein
PF00833	Ribosomal_S17e	24.2	Ribosomal S17
PF00861	Ribosomal_L18p	51.85	Ribosomal L18p/L5e family
PF00867	XPG_I	39.9	XPG I-region
PF00900	Ribosomal_S4e	40.7	Ribosomal family S4e
PF00919	UPF0004	31.45	Uncharacterized protein family UPF0004
PF00935	Ribosomal_L44	41.9	Ribosomal protein L44
PF00958	GMP_synt_C	47.05	GMP synthase C terminal domain
PF01000	RNA_pol_A_bac	40.2	RNA polymerase Rpb3/RpoA insert domain
PF01015	Ribosomal_S3Ae	111.55	Ribosomal S3Ae family
PF01035	DNA_binding_1	38.1	6-O-methylguanine DNA methyltransferase, DNA binding domain
PF01090	Ribosomal_S19e	84.1	Ribosomal protein S19e
PF01092	Ribosomal_S6e	48.7	Ribosomal protein S6e
PF01139	UPF0027	244	Uncharacterized protein family UPF0027
PF01157	Ribosomal_L21e	40.35	Ribosomal protein L21e
PF01172	SBDS	38.7	Shwachman-Bodian-Diamond syndrome (SBDS) protein
PF01174	SNO	82.5	SNO glutamine amidotransferase family
PF01180	DHO_dh	71.95	Dihydroorotate dehydrogenase
PF01191	RNA_pol_Rpb5_C	41.75	RNA polymerase Rpb5, C-terminal domain
PF01192	RNA_pol_Rpb6	23.5	RNA polymerase Rpb6
PF01194	RNA_pol_N	41.25	RNA polymerases N / 8 kDa subunit
PF01198	Ribosomal_L31e	35.6	Ribosomal protein L31e
PF01200	Ribosomal_S28e	44.25	Ribosomal protein S28e
PF01201	Ribosomal_S8e	53.6	Ribosomal protein S8e
PF01246	Ribosomal_L24e	30.9	Ribosomal protein L24e
PF01253	SUI1	34.75	Translation initiation factor SUI1

PF01269	Fibrillarin	143.65	Fibrillarin
PF01280	Ribosomal_L19e	85.95	Ribosomal protein L19e
PF01282	Ribosomal_S24e	36	Ribosomal protein S24e
PF01287	eIF-5a	20.65	Eukaryotic elongation factor 5A hypusine, DNA-binding OB fold
PF01351	RNase_HII	67.7	Ribonuclease HII
PF01496	V_ATPase_I	85.05	V-type ATPase 116kDa subunit family
PF01509	TruB_N	31.5	TruB family pseudouridylate synthase (N terminal domain)
PF01599	Ribosomal_S27	28.6	Ribosomal protein S27a
PF01655	Ribosomal_L32e	57.3	Ribosomal protein L32
PF01667	Ribosomal_S27e	31	Ribosomal protein S27
PF01725	Ham1p_like	78.25	Ham1 family
PF01780	Ribosomal_L37ae	46.15	Ribosomal L37ae protein family
PF01798	Nop	80.45	Putative snoRNA binding domain
PF01813	ATP-synt_D	94.35	ATP synthase subunit D
PF01849	NAC	24.05	NAC domain
PF01864	DUF46	76.25	Putative integral membrane protein DUF46
PF01866	Diphthamide_syn	149.75	Putative diphthamide synthesis protein
PF01868	UPF0086	37.8	Domain of unknown function UPF0086
PF01884	PcrB	47.8	PcrB family
PF01896	DNA_primase_S	25.85	Eukaryotic and archaeal DNA primase small subunit
PF01907	Ribosomal_L37e	34.15	Ribosomal protein L37e
PF01912	eIF-6	104.05	eIF-6 family
PF01920	Prefoldin_2	34.45	Prefoldin subunit
PF01922	SRP19	35.95	SRP19 protein
PF01941	AdoMet_Synthase	285.4	S-adenosylmethionine synthetase (AdoMet synthetase)
PF01948	PyrI	54.45	Aspartate carbamoyltransferase regulatory chain, allosteric domain
PF01949	DUF99	92.65	Protein of unknown function DUF99
PF01951	Archease	62.7	Archease protein family (MTH1598/TM1083)
PF01967	MoaC	75	MoaC family
PF01973	MAF_flag10	46.8	Protein of unknown function DUF115
PF01981	PTH2	64.6	Peptidyl-tRNA hydrolase PTH2
PF01982	CTP-dep_RFKase	62.9	Domain of unknown function DUF120
PF01984	dsDNA_bind	43.1	Double-stranded DNA-binding domain
PF01990	ATP-synt_F	26.2	ATP synthase (F/14-kDa) subunit
PF01991	vATP-synt_E	55.45	ATP synthase (E/31 kDa) subunit
PF02005	TRM	144.1	N2,N2-dimethylguanosine tRNA methyltransferase
PF02006	DUF137	119.55	Protein of unknown function DUF137
PF02748	PyrI_C	25.95	Aspartate carbamoyltransferase regulatory chain, metal binding domain
PF02834	LigT_PEase	40.65	LigT like Phosphoesterase
PF02938	GAD	22.75	GAD domain
PF02978	SRP_SPB	46.7	Signal peptide binding domain
PF02996	Prefoldin	34.95	Prefoldin subunit
PF03439	Spt5-NGN	32.45	Early transcription elongation factor of RNA pol II, NGN section
PF03484	B5	25.45	tRNA synthetase B5 domain

PF03719	Ribosomal_S5_C	35.45	Ribosomal protein S5, C-terminal domain
PF03764	EFG_IV	46.75	Elongation factor G, domain IV
PF03853	YjeF_N	68.05	YjeF-related protein N-terminus
PF03874	RNA_pol_Rpb4	17.45	RNA polymerase Rpb4
PF03876	SHS2_Rpb7-N	25.4	SHS2 domain found in N terminus of Rpb7p/Rpc25p/MJ0397
PF03911	Sec61_beta	15.55	Sec61beta family
PF03946	Ribosomal_L11_N	35.5	Ribosomal protein L11, N-terminal domain
PF03947	Ribosomal_L2_C	67.75	Ribosomal Proteins L2, C-terminal domain
PF03950	tRNA-synt_1c_C	37.35	tRNA synthetases class I (E and Q), anti-codon binding domain
PF04010	DUF357	32.85	Protein of unknown function (DUF357)
PF04019	DUF359	56.65	Protein of unknown function (DUF359)
PF04104	DNA_primase_lrg	51.55	Eukaryotic and archaeal DNA primase, large subunit
PF04127	DFP	79.55	DNA / pantothenate metabolism flavoprotein
PF04135	Nop10p	35.05	Nucleolar RNA-binding protein, Nop10p family
PF04406	TP6A_N	32.45	Type IIB DNA topoisomerase
PF04414	tRNA_deacylase	104.55	D-aminoacyl-tRNA deacylase
PF04560	RNA_pol_Rpb2_7	42.4	RNA polymerase Rpb2, domain 7
PF04561	RNA_pol_Rpb2_2	45.9	RNA polymerase Rpb2, domain 2
PF04563	RNA_pol_Rpb2_1	89.55	RNA polymerase beta subunit
PF04565	RNA_pol_Rpb2_3	39.65	RNA polymerase Rpb2, domain 3
PF04566	RNA_pol_Rpb2_4	32.5	RNA polymerase Rpb2, domain 4
PF04567	RNA_pol_Rpb2_5	21.4	RNA polymerase Rpb2, domain 5
PF04919	DUF655	104.9	Protein of unknown function, DUF655
PF04981	NMD3	84.6	NMD3 family
PF04983	RNA_pol_Rpb1_3	53.05	RNA polymerase Rpb1, domain 3
PF04997	RNA_pol_Rpb1_1	142.85	RNA polymerase Rpb1, domain 1
PF05000	RNA_pol_Rpb1_4	48.5	RNA polymerase Rpb1, domain 4
PF05221	AdoHcyase	105.15	S-adenosyl-L-homocysteine hydrolase
PF05670	DUF814	39.6	Domain of unknown function (DUF814)
PF05746	DALR_1	46.9	DALR anticodon binding domain
PF06026	Rib_5-P_isom_A	88.6	Ribose 5-phosphate isomerase A (phosphoriboisomerase A)
PF06093	Spt4	18.4	Spt4/RpoE2 zinc finger
PF06418	CTP_synth_N	205	CTP synthase N-terminus
PF06463	Mob_synth_C	40.05	Molybdenum Cofactor Synthesis C
PF07541	EIF_2_alpha	41.75	Eukaryotic translation initiation factor 2 alpha subunit
PF08068	DKCLD	28.05	DKCLD (NUC011) domain
PF08069	Ribosomal_S13_N	36.15	Ribosomal S13/S15 N-terminal domain
PF08071	RS4NT	21.25	RS4NT (NUC023) domain
PF08608	Wyosine_form	30.8	Wyosine base formation
PF08704	GCD14	38.5	tRNA methyltransferase complex GCD14 subunit
PF09173	eIF2_C	46.35	Initiation factor eIF2 gamma, C terminal
PF09239	Topo-VIb_trans	100.4	Topoisomerase VI B subunit, transducer
PF09249	tRNA_NucTransf2	66.75	tRNA nucleotidyltransferase, second domain
PF09377	SBDS_C	31.2	SBDS protein C-terminal domain

PF10559	Plug_translocon	14.9	Plug domain of Sec61p
PF11987	IF-2	34.05	Translation-initiation factor 2

Supplementary Table 14 | Genes used to calculate the Patescibacteria versus Terrabacteria tree. The resulting phylogentic tree is shown in Supplementary Figure 9.

HMM name	Gene name	HMM length
PF00162	Phosphoglycerate kinase	384
PF00276	Ribosomal protein L23	92
PF00281	Ribosomal protein L5	56
PF00297	Ribosomal protein L3	263
PF00380	Ribosomal protein S9/S16	121
PF00410	Ribosomal protein S8	129
PF00411	Ribosomal protein S11	110
PF00416	Ribosomal protein S13/S18	106
PF00466	Ribosomal protein L10	100
PF00573	Ribosomal protein L4/L1 family	192
PF01795	MraW methylase family	310
TIGR00001	rpml_bact: ribosomal protein L35	63
TIGR00002	S16: ribosomal protein S16	78
TIGR00019	prfA: peptide chain release factor 1	361
TIGR00029	S20: ribosomal protein S20	87
TIGR00043	TIGR00043: probable rRNA maturation factor YbeY	111
TIGR00059	L17: ribosomal protein L17	112
TIGR00060	L18_bact: ribosomal protein L18	114
TIGR00061	L21: ribosomal protein L21	101
TIGR00064	ftsY: signal recognition particle-docking protein FtsY	279
TIGR00082	rbfA: ribosome-binding factor A	115
TIGR00086	smpB: SsrA-binding protein	144
TIGR00092	TIGR00092: GTP-binding protein YchF	368
TIGR00115	tig: trigger factor	410
TIGR00116	tsf: translation elongation factor Ts	293
TIGR00158	L9: ribosomal protein L9	148
TIGR00165	S18: ribosomal protein S18	70
TIGR00166	S6: ribosomal protein S6	95
TIGR00168	infC: translation initiation factor IF-3	165
TIGR00337	PyrG: CTP synthase	526
TIGR00344	alaS: alanine--tRNA ligase	847
TIGR00362	DnaA: chromosomal replication initiator protein DnaA	437
TIGR00388	glyQ: glycine--tRNA ligase, alpha subunit	293
TIGR00389	glyS_dimeric: glycine--tRNA ligase	565
TIGR00409	proS_fam_II: proline--tRNA ligase	568
TIGR00459	aspS_bact: aspartate--tRNA ligase	586
TIGR00460	fmt: methionyl-tRNA formyltransferase	315
TIGR00468	pheS: phenylalanine--tRNA ligase, alpha subunit	324
TIGR00471	pheT_arch: phenylalanine--tRNA ligase, beta subunit	551
TIGR00472	pheT_bact: phenylalanine--tRNA ligase, beta subunit	798
TIGR00487	IF-2: translation initiation factor IF-2	587
TIGR00496	frr: ribosome recycling factor	176
TIGR00575	dnlj: DNA ligase, NAD-dependent	652
TIGR00631	uvrb: excinuclease ABC subunit B	658
TIGR00663	dnan: DNA polymerase III, beta subunit	367

TIGR00810	secG: preprotein translocase, SecG subunit	73
TIGR00855	L12: ribosomal protein L7/L12	125
TIGR00922	nusG: transcription termination/antitermination factor NusG	172
TIGR00959	ffh: signal recognition particle protein	428
TIGR00963	secA: preprotein translocase, SecA subunit	787
TIGR00964	secE_bact: preprotein translocase, SecE subunit	57
TIGR00967	3a0501s007: preprotein translocase, SecY subunit	414
TIGR00981	rpsL_bact: ribosomal protein S12	124
TIGR01009	rpsC_bact: ribosomal protein S3	212
TIGR01011	rpsB_bact: ribosomal protein S2	225
TIGR01021	rpsE_bact: ribosomal protein S5	156
TIGR01024	rplS_bact: ribosomal protein L19	114
TIGR01029	rpsG_bact: ribosomal protein S7	154
TIGR01032	rplT_bact: ribosomal protein L20	114
TIGR01044	rplV_bact: ribosomal protein L22	103
TIGR01049	rpsJ_bact: ribosomal protein S10	99
TIGR01050	rpsS_bact: ribosomal protein S19	92
TIGR01063	gyrA: DNA gyrase, A subunit	800
TIGR01066	rplM_bact: ribosomal protein L13	141
TIGR01067	rplN_bact: ribosomal protein L14	122
TIGR01071	rplO_bact: ribosomal protein L15	144
TIGR01079	rplX_bact: ribosomal protein L24	104
TIGR01164	rplP_bact: ribosomal protein L16	126
TIGR01169	rplA_bact: ribosomal protein L1	227
TIGR01171	rplB_bact: ribosomal protein L2	275
TIGR01391	dnaG: DNA primase	414
TIGR01393	lepA: GTP-binding protein LepA	595
TIGR01632	L11_bact: ribosomal protein L11	140
TIGR01953	NusA: transcription termination factor NusA	340
TIGR02012	tigrfam_recA: protein RecA	321
TIGR02013	rpoB: DNA-directed RNA polymerase, beta subunit	1238
TIGR02027	rpoA: DNA-directed RNA polymerase, alpha subunit	298
TIGR02386	rpoC_TIGR: DNA-directed RNA polymerase, beta' subunit	1147
TIGR02387	rpoC1_cyan: DNA-directed RNA polymerase, gamma subunit	619
TIGR02397	dnaX_nterm: DNA polymerase III, subunit gamma and tau	355
TIGR02729	Obg_CgtA: Obg family GTPase CgtA	329
TIGR03263	guanylyl_kin: guanylate kinase	180
TIGR03594	GTPase_EngA: ribosome-associated GTPase EngA	432

Supplementary Table 15 | COGs and alignment lengths used to calculate gene trees (see Supplementary methods).

COG id	gene name	masked alignment position/ length (aa)
COG1252	oxidoreductase	453-1129/676
COG1759	PurP	75-485/410
COG1595	sigma factor	50-125/75; 220-280/60
COG0317	SpoT	80-527/177
COG0741	murein transglycosylase	1527-1852/325